Journal of
Clinical Psychology

JANUARY 1954
Volume X Number 1

EDITORIAL BOARD

FREDERICK C. THORNE, *Editor*
Medical College, University of Vermont

JERRY W. CARTER, JR.
U. S. Public Health Service

HOWARD F. HUNT
University of Chicago

WILLIAM A. HUNT
Northwestern University

ELAINE F. KINDER
New York State Department of Mental Hygiene

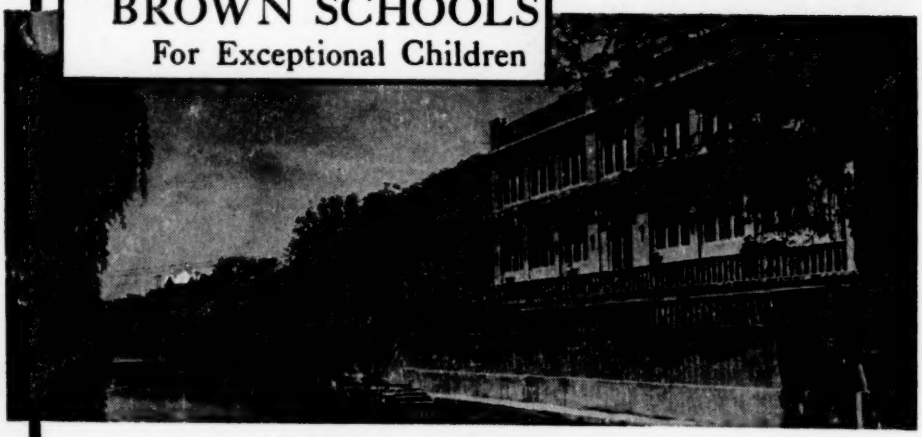
C. M. LOUTTIT
University of Illinois

ROBERT I. WATSON
Northwestern University

ELEANOR C. THORNE, *Business Manager*

Journal of Clinical Psychology
5 Pearl Street · · Brandon, Vermont

THE BROWN SCHOOLS
For Exceptional Children



YEAR-ROUND SCHOOL - SUMMER CAMP PROGRAM
SIX RESIDENCE CENTERS

Exclusive private school offering a specialized program of education and re-education for children with educational and emotional difficulties. Members of the allied professions are invited to write for descriptive literature and to confer with our experienced staff in individual cases.

- Suburban and Ranch Locations
- Ideal Year-Round Climate
- Academic Instruction
- Handicraft Classes
- Home for Older Girls
- Lakeside Lodges for Summer Camp
- Resident Registered Nurses

BERT P. BROWN, F. A. A. M. D., *President*
LYNDON L. BROWN, *Vice President*
PAUL L. WHITE, M. D., F. A. C. P., *Medical Director*
M. D. HEATTY, M. D., F. A. C. S., *Resident Physician*
JESSE VILLAREAL, Ph.D., *Speech Pathologist*
JEAN GIESEY MIMS, M. A., *Clinical Psychologist*

P.O. Box 4008-D
AUSTIN, TEXAS

WRITE TO NOVA LEE DEARING, *Registrar*

Interviews arranged with the Consulting Staff at the School or in the Home

JOURNAL OF CLINICAL PSYCHOLOGY

VOL. X JANUARY 1954 No. 1

CONTENTS

Page

Pattern Analysis and Configural Scoring. Paul Horst, University of Washington 3
A Factor Analysis of Curvilinear Distortions on the Bender-Gestalt. Wilson H. Guertin, Veterans Administration Hospital, Knoxville, Iowa 12
A Story Completion Test for College Students. Eugene S. Mills, Whittier College 18
Personality Integration and the Perceptual Process. Eli A. Rubinstein, Veterans Administration, Washington, D. C. 23
An Analysis of the Rationalization and Explanations of Test Behavior. Joseph E. Shorr, Los Angeles, Calif. 29
A Quantitative Study of the HTP and its Relationship to the Wechsler-Bellevue Scale. Harold Rubin, Veterans Administration Hospital, Lebanon, Penna. 35
The Use of Cartoons as a Projective Device. George B. Strother, Margaret McCrackin Barnett and Peter C. Apostolakos, University of Minnesota, Duluth Branch 38
A Preliminary Report on a Technique Designed to Differentiate Patients with Cerebral Pathology and with Psychoneurosis. Alexander Tolor, Neurological Institute, Presbyterian Hospital, N. Y. 43
A Preliminary Study to Develop a More Discriminating F+ Ratio. Marvin J. Feldman, Carolyn Gursslin, Marvin L. Kaplan and Nidia Sharlock, University of Buffalo 47
The Effects of Recent Perceptual Training and Experience on Rorschach Performance. Irwin J. Knopf, State University of Iowa 52
Contributions to a Theory of Selecting Psychotherapy Patients. Frank Auld, Jr. and Jerome K. Myers, Yale University 56
A Study of Twelve-Year-Old Recidivists. William W. Wattenberg and Frank Quiroz. Wayne University 61
The Use of Drawings of the Human Figure as an Adjunct in Psychotherapy. Francis W. King, Dartmouth College 65
Self-Perceptions in Stuttering. Seymour L. Zelen, Joseph G. Sheehan and James F. T. Bugental. University of California, Los Angeles 70
The Diagnostic Value of the Rorschach P-F Study. Robert L. Brown, Furman University and Oliver L. Lacey, University of Alabama 72
Case Report on a Mental Hygiene Clinic Patient Tested Before and After Frontal Lobe Injury. Lawrence S. Rogers and James W. Taylor, Veterans Administration, Denver, Colorado 75
Rorschach Pattern Analysis in Schizophrenia: A Cross-Validation Study. Earl S. Taulbee, Norfolk (Nebraska) State Hospital and Boyd D. Sisson, University of Nebraska 80
A Mechanical Short-Cut for Computing Tetrachoric Correlations. Anders Sweetland and Ralph Mason Dreger, Florida State University. 82

(Over)

CONTENTS—Continued

The Learning Plateau Seen as Resistance in Individual Remedial Teaching. Monroe G. Gottsegen, Inter Agency Guidance Center, Yonkers, N. Y. and Gloria B. Gottsegen The Developmental School, New York, N. Y. 85
Concept Formation of Normal and Subnormal Adolescents on a Modification of the Weigl-Goldstein-Scheerer Color Form Sorting Test. Arne Korstvedt, Chalmers L. Stacey and William F. Reynolds. Syracuse University. 88
Hidden Costs in the Utilization of the Psychiatrically Marginal Man. William A. Hunt, Northwestern University, Cecil L. Wittson, University of Nebraska College of Medicine, and Edna B. Hunt 91
Correlates of the Non-Committal Test-Item Response. Grace Rubin-Rabson, Fort Wayne, Indiana. 93
The Relationship between Z Scores on the Bender Gestalt and F+% on the Rorschach. Robert H. Curnutt and William B. Lewis, U. S. Naval Hospital, National Naval Medical Center. 96
Correlation of Clinical Estimates with Test Scores on Mental Ability and Personality Tests. J. Watson Wilson, Nordli, Ogan, Wilson Associates, New York, N. Y. 97

Editorial Opinions 100
Book Notices 103

GENERAL INFORMATION

The *Journal of Clinical Psychology* is an independent journal dedicated to the advancement of the clinical method in psychology. Although primarily a scientifically oriented professional journal limited to publication of original research reports and authoritative theoretical articles, it aims to foster the promotion and expansion of clinical psychology as an applied science. Special sections in each issue will be devoted to news, comments, and letters to the editors on matters pertinent to the profession. Communications and articles submitted for publication are welcome and should be addressed to the Editorial Office, 5 Pearl Street, Brandon, Vermont.

ORIGINAL ARTICLES are published only with the understanding that they are contributed exclusively to the *Journal of Clinical Psychology*. The publishers are not responsible for statements made or opinions expressed by contributors in articles printed in these columns. Copyrights cover publication in the *Journal of Clinical Psychology* and articles may not be reproduced without permission of the publishers.

MANUSCRIPTS must be typewritten, double-spaced with good margins, on one side of the paper only. The publishers will not be responsible for manuscripts lost in the mails or through unavoidable accident. Retain a carbon copy after sending the original to the editor. It is requested that all articles be as brief as consistent with clarity. All manuscripts are subject to editorial modification.

THE STYLE of all manuscripts including bibliographies must be in accordance with the instructions given in: Anderson, J. E., and Valentine, W. L. The preparation of articles for publication in the journals of the American Psychological Association. *Psychol. Bull.*, 1944, 41, 345-376. References should be arranged alphabetically according to the directions given above. Full address of the author should appear somewhere in the article, preferably at the end.

REPRINTS are furnished on order only and must be requested on the order blank provided when galley proofs are returned. Change of address notices should be sent one month before moving and should include old and new addresses. Undelivered copies will be remailed only at subscribers' expense.

The *Journal of Clinical Psychology* is published quarterly, appearing in January, April, July and October, at Rutland, Vermont, by Frederick C. Thorne, M.D. Subscription price $7.50; Canadian subscriptions $8.00; foreign subscriptions $8.50 including postage; single copies $2.00. Entered as second class matter January 15, 1945, at the post office at Burlington, Vermont, under the Act of March 3, 1879. Reentered at the post office at Brandon, Vermont. Additional entry at the post office at Rutland, Vermont. *Title registered, U. S. Patent Office.* Editorial and Business Offices, 5 Pearl Street, Brandon, Vermont.

Copyright, 1954, Frederick C. Thorne, M. D.

PATTERN ANALYSIS AND CONFIGURAL SCORING

PAUL HORST

University of Washington

Problem

Recently Meehl[2] has called attention to the emphasis which clinicians place on the pattern of a client's responses. In particular he considers the concept of patterning with respect to a person's responses to the items in a test. He cites as an example two hypothetical dichotomously scored test items each of which has a phi-coefficient of zero with a dichotomous criterion such as a "schizophrenic *vs* normal" group. Numerical examples are provided to show that even with a zero validity for each item, it may still be possible to differentiate the two criterion groups on the basis of the responses to the two items. To achieve this differentiation, a person is credited with a correct response if both items are answered in the same way (either true or false), and with an incorrect response if they are answered differently. Meehl designates this type of scoring as "configural scoring", and suggests that it may provide a basis for developing predictive instruments even though the separate items in the instrument may have negligible validity. This approach seems to offer the possibility of objective qualification to those elusive patterns of response which clinicians have regarded as important.

On the assumption that each of a pair of items correlated zero with a dichotomous criterion, Meehl was concerned with the following question: "Having given the inter-item phi-coefficient for each of the two criterion groups then, if the items were configurally scored, what would be the correlation of this score with the dichotomous criterion?" He makes the assumption that the proportion of cases in the two criterion groups is fifty per cent, and that within each criterion group the proportion answering each item "true" is also fifty per cent. He then determines empirically the configural validity of the items for certain specified values of the inter-item phi-coefficients in each of the two criterion groups.

The purpose of this article is twofold. One is to show the precise relationship between "configural" validity of the items. The second is to point out that the conventional method of scoring a test consists of a linear combination of item scores, whereas the "configural" method proposed by Meehl is a special case of a non-linear combination of item scores.

Mathematical Rationale

Suppose we take the special case considered by Meehl where the two criterion groups are of the same size and fifty per cent of each criterion group mark each of the two items "true." We let

Φ_1 = the phi-coefficient for the two items for sub group 1, e.g. the normal group

Φ_2 = the phi-coefficient for the two items for sub group 2, e.g. the schizophrenic group

Φ_T = the phi-coefficient for the configural score on the two items with the dichotomous criterion

Following Meehl, the configural score on the two items is "1" if both items are marked the same either "true "or "false" and "O" if either one of the items is marked "true" and the other is marked "false."

It can be shown then* that Φ_T is given by the equation.

$$\Phi_T = \frac{\Phi_1 - \Phi_2}{\sqrt{4 - (\Phi_1 + \Phi_2)^2}} \qquad (1)$$

*See appendix for proof.

This formula explains rather simply the phenomena noted by Meehl with reference to the configural validity of two items. He found empirically that if $\Phi_2 = -\Phi_1$, that is if Φ_2 is equal in absolute value but opposite in sign to Φ_1, then the configural validity would be equal to Φ_1. The reason for this is evident from the formula.

If $\Phi_2 = -\Phi_1$ we have from equation (1)

$$\Phi_T = \frac{\Phi_1 - (-\Phi_1)}{\sqrt{4 - (\Phi_1 - \Phi_1)^2}}$$

or

$$\Phi_T = \frac{2\Phi_1}{\sqrt{4}}$$

or simply

$$\Phi_T = \Phi_1 \qquad (2)$$

Meehl also found empirically that $\frac{\Phi_1 - \Phi_2}{2}$ gives a good estimate of Φ_T provided the average or sum of Φ_1 and Φ_2 is not too large. Formula (1) shows clearly why this must be true. As $\Phi_1 + \Phi_2$ becomes larger the denominator term becomes smaller. Therefore, the number by which $\Phi_1 - \Phi_2$ must be divided becomes increasingly less than 2. It is also obvious from formula (1) that when $\Phi_1 = \Phi_2$ then Φ_T must be zero.

Table 1 gives the value of the configural validity for specified values of Φ_1 and Φ_2.

TABLE 1. VALUES OF THE CONFIGURAL VALIDITY Φ_T CORRESPONDING TO SPECIFIED VALUES OF Φ_1 AND Φ_2

Φ_2 \ Φ_1	−1.0	−.8	−.6	−.4	−.2	0	.2	.4	.6	.8	1.0
−1.0	0	.229	.333	.420	.500	.577	.655	.734	.816	.905	1.000
− .8	−.229	0	.140	.250	.346	.437	.524	.612	.704	.800	.905
− .6	−.333	−.140	0	.115	.218	.314	.408	.503	.600	.704	.816
− .4	−.420	−.250	−.115	0	.105	.204	.302	.400	.503	.612	.734
− .2	−.500	−.346	−.218	−.105	0	.100	.200	.302	.408	.524	.655
0	−.577	−.437	−.314	−.204	−.100	0	1.00	.204	.314	.437	.577
.2	−.655	−.524	−.408	−.302	−.200	−.100	0	.105	.218	.346	.500
.4	−.734	−.612	−.503	−.400	−.302	−.204	−.105	0	.115	.250	.420
.6	−.816	−.704	−.600	−.503	−.408	−.314	−.218	−.115	0	.140	.333
.8	−.905	−.800	−.704	−.612	−.524	−.437	−.346	−.250	−.140	0	.229
1.0	−1.000	−.905	−.816	−.734	−.655	−.577	−.500	−.420	−.333	−.229	0

We arbitrarily let columns of the table represent different values of Φ_1 ranging from −1 to +1. The rows then represent values of Φ_2 ranging from −1 to +1. For example, suppose we wish to find the configural validity corresponding to a Φ_1 of .4 and a Φ_2 of −.2. We look for the entry which lies in the column headed .4 and the row with −.2 at its left. This entry is .302.

Figure 1 shows graphically the relationship given in the tables. Each curve represents for a specified value of Φ_2 the relationship between Φ_T and Φ_1. As Φ_2

Relationship between the configural validity of two test items as a function of their ϕ coefficients. (See text)

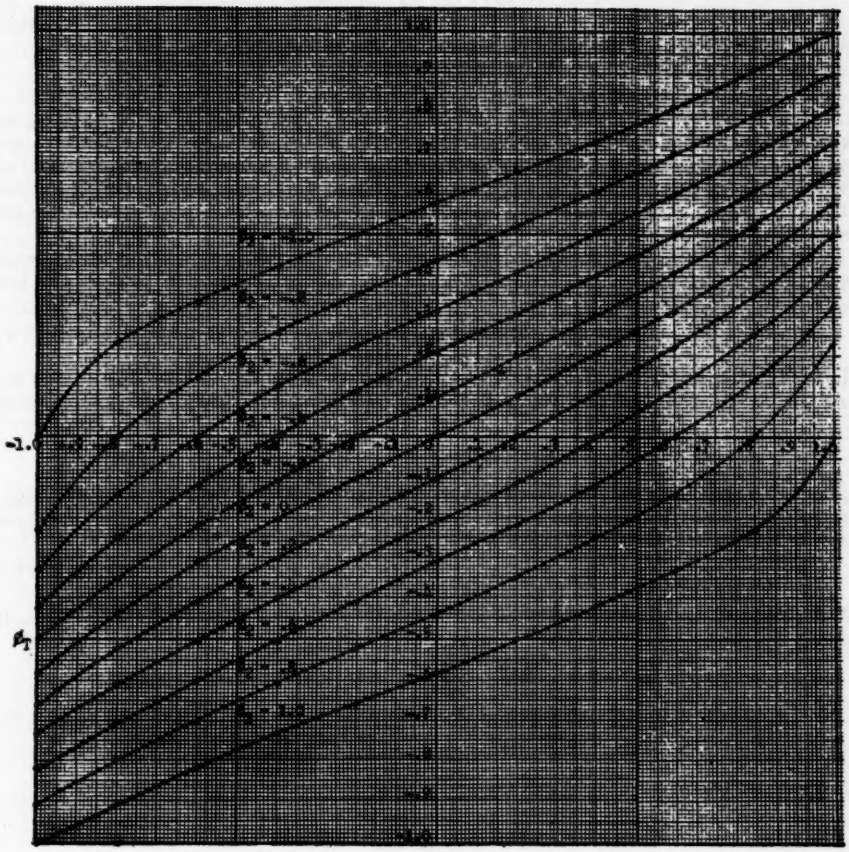


Figure 1

approaches zero the relationship between Φ_T and Φ_1 becomes more nearly linear although even for $\Phi_2 = 0$, this relationship is not a straight line. It is also clear from the chart that as Φ_2 departs either positively or negatively from 0 the departure of the curves from linearity are greatest as Φ_1 approaches Φ_2. For example, when $\Phi_2 = -1.0$, the departure of the curve from linearity becomes great only as Φ_1 is close to -1.0. The same is true for $\Phi_2 = +1.0$ and Φ_1 close to $+1.0$. This is, of course, the phenomenon which Meehl discovered empirically.

Next let us see how we may regard "configural" scoring as a special case of a non linear combination of item responses. We start with the assumption that a score on a single item is either "1" or "0" depending on which of two ways it was answered. We shall not here consider items having more than two possible responses or item scoring procedures admitting of more then two different scores for an item. The principle outlined here can be generalized to such cases, but it is beyond the scope of this paper to do so.

Now the usual procedure for getting a total score from a group of items is simply to add the number of item scores, that is, the number of items answered "correctly." First let us note that this procedure is but a special case of the more general linear combination of measures. If we have for a person a set of test scores ($x_1, x_2; \ldots x_n$) on n different tests then we may get a linear combination T of these scores by assigning a weight to each score so that

$$T = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n \qquad (3)$$

where b_0 is an additive constant. In particular the b's in equation (3) may be regression weights, determined in such a way as to give the best prediction of some criterion measure. But suppose now that we let b_0 in equation (3) be "0" and all the other b's we take as "1". Suppose also that each test in equation (3) consists only of a single item in which the score is either "0" or "1". Then T becomes the conventional test score consisting of number of items correct.

But except for the sake of simplicity, there is no particular reason why we must get a total score based on a linear combination of measures. If we have only two tests, or items, for example, we could get a total score, as shown in equation (4).

$$T = b_0 + b_1 x_1 + b_2 x_2 + b_{12} x_1 x_2 \qquad (4)$$

Here the total score is a linear combination of the scores on x_1 and x_2 and also the product of the scores, $x_1 x_2$. Suppose now we let $b_0 = 1$, $b_1 = -1$, $b_2 = -1$, $b_{12} = 2$. Then equation (4) would be

$$T = 1 - x_1 - x_2 + 2x_1 x_2 \qquad (5)$$

Now suppose that tests 1 and 2 are single items with possible scores of only "1" and "0". If a person gets a score of "1" on both items, then $x_1 = 1$, $x_2 = 1$ and $x_1 x_2 = 1 \times 1 = 1$. We would have in equation 5, therefore

$$T = 1 - 1 - 1 + 2 = 1$$

Next suppose a person gets a score of "0" on both items. We would have $x_1 = 0$, $x_2 = 0$, $x_1 x_2 = 0 \times 0 = 0$ so that for equation (5) we get

$$T = 1 - 0 - 0 - 2 \times 0 = 1$$

Now if a person gets "0" on item 1 and "1" on item 2, we would have in equation (5)

$$T = 1 - 0 - 1 + 2 \times 1 \times 0 = 0$$

Similarly, if he gets "1" on item 1 and "0" on item 2, we would have

$$T = 1 - 1 - 0 + 2 \times 0 \times 1 = 0$$

We see therefore that equation (5) is the precise mathematical formula for the special case of configural scoring described by Meehl.* It is also of interest to note that equation (5) can be written in the form

$$T = 1 - (x_1 - x_2)^2 \qquad (6)$$

or what amounts to the same thing

$$T = 1 - (x_2 - x_1)^2 \qquad (7)$$

From either equation (6) or (7) it is easy to see that if both items have the same score then the term in parenthesis must be zero and T is "1". If on the other hand one score is "0" and the other "1", then the term in the parenthesis is either "+1" or "−1" depending on which score is "0". But in either case the square of this difference is "1" which makes T equal to zero.

*See appendix for proof that the weights used in equation (5) result in the maximum correlation of T with the criterion.

Configural Scoring by the "H" Technique

Another interesting example of a type of configural scoring of items is given by Stouffer, *et al.*[3] They have developed a procedure for selecting and scoring a set of items which they have named "H technique." By means not relevant to our discussion triads or groups of 3 items per group are selected. Each triad is then regarded as a single item. The methods for selecting and scoring the triads purport to yield item triads for a scale which give a higher coefficient of reproducibility than is the case when single items are used and scored dichotomously. Extension and generalization of the procedure seems to have promise not only for clinical psychology but also for other areas concerned with psychological measurement and evaluation. Of particular interest in our discussion is the method of scoring the triads specified by Stouffer et al. The procedure is simply to assign a score of "1" if all or any two items in the triad are "correct" and "0" if none or only one are "correct." This again is an example of what we may call "configural" scoring. But let us see how we may express such a score as a special case of a non linear combination of measures. Suppose instead of a linear combination of measures given by equation (4) we consider for the case of 3 tests, a total score which involves also all possible products of scores. Such a score would be given by equation (8).

$$T = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{123}x_1x_2x_3 \quad (8)$$

Now suppose we let $b_0 = b_1 = b_2 = b_3 = 0$, $b_{12} = b_{13} = b_{23} = 1$, and $b_{123} = -2$. Then equation (8) becomes

$$T = x_1x_2 + x_1x_3 + x_2x_3 - 2x_1x_2x_3 \quad (9)$$

It is easy to show that equation (9) is the precise mathematical scoring formula for the triads which gives a score of "1" if all or any two items in the triad are "correct" and "0" if none or only one item in the triad are "correct". Consider the various possible patterns of item scores for a triad below.

Pattern	Item No. 1	2	3	Triad score
1	1	1	1	1
2	1	1	0	1
3	1	0	1	1
4	0	1	1	1
5	1	0	0	0
6	0	1	0	0
7	0	0	1	0
8	0	0	0	0

If the x values in equation (9) are taken in turn from the successive patterns of item scores given in the table above, we have the following solutions for the equation.

Pattern No.

1 $T = 1 \times 1 + 1 \times 1 + 1 \times 1 - 2 \times 1 \times 1 \times 1 = 1$
2 $T = 1 \times 1 + 1 \times 0 + 1 \times 0 - 2 \times 1 \times 1 \times 0 = 1$
3 $T = 1 \times 0 + 1 \times 1 + 0 \times 1 - 2 \times 1 \times 0 \times 1 = 1$
4 $T = 0 \times 1 + 0 \times 1 + 1 \times 1 - 2 \times 0 \times 1 \times 1 = 1$
5 $T = 1 \times 0 + 1 \times 0 + 0 \times 0 - 2 \times 1 \times 0 \times 0 = 0$
6 $T = 0 \times 1 + 0 \times 0 + 1 \times 0 - 2 \times 0 \times 1 \times 0 = 0$
7 $T = 0 \times 0 + 0 \times 1 + 0 \times 1 - 2 \times 0 \times 0 \times 1 = 0$
8 $T = 0 \times 0 + 0 \times 0 + 0 \times 0 - 2 \times 0 \times 0 \times 0 = 0$

Although we have given only two specific examples of how "configural" scoring techniques can be expressed in terms of precise mathematical formulas the methods may be greatly extended and generalized to cover almost any configural scoring pro-

cedure which might be proposed. No doubt, with good reason, there is currently a great deal of interest and activity in the analysis of patterns and configurations of responses as indicated by the review of Gaier and Lee [1]. It is highly probable, however, that the work in this area will be much more fruitful when more precise and rigorous mathematical concepts are developed to take the place of verbal formulations and analyses based on empirical or trial and error manipulations of the data.

Summary

A mathematical formula is presented showing the precise relationship between the "configural" validity of two test items and the inter-item phi-coefficients for each of the two subgroups of a dichotomous criterion. The formula predicts the phenomena discussed by Meehl [2]. Two examples are taken, one from Meehl and one from Stouffer [3] to show how exact mathematical formulas may be used to express a "configural" score. It is suggested that tbe analysis of patterns or configurations of responses may become much more fruitful with the introduction of non linear mathematical formulations and procedures.

Appendix: Proof of Equation 1 and 5.

Suppose we have two dichotomous independent variables, X and Y, and one dichotomous dependent variable, Z. Consider also a third independent variable which is the product XY. All correlation coefficients among the variables are to be taken as product moment correlations and since all variables are dichotomous they will be phi-coefficients. We let

$$\Phi = \begin{bmatrix} 1 & \Phi_{XY} & \Phi_{X(XY)} \\ \Phi_{XY} & 1 & \Phi_{Y(XY)} \\ \Phi_{X(XY)} & \Phi_{Y(XY)} & 1 \end{bmatrix} \tag{1}$$

$$\Phi_z = \begin{bmatrix} \Phi_{XZ} \\ \Phi_{YZ} \\ \Phi_{(XY)Z} \end{bmatrix} \tag{2}$$

where the Φ's are phi coefficients between the variables indicated by the subscripts. We assume now that

$$\Phi_{XZ} = \Phi_{YZ} = 0 \tag{3}$$

$$M_X = M_Y = M_Z = 1/2 \tag{4}$$

Because of 4 it can be proved that

$$\Phi_{Y(XY)} = \Phi_{X(XY)} \tag{5}$$

If we let Φ_T be the multiple correlation of the variable Z with X, Y, and XY, it is well known that

$$\Phi_T^2 = \Phi_z^1 \, \Phi^{-1} \Phi_z \tag{6}$$

Because of 3, 5 and 6 we can show that in scalar notation

$$\Phi_T^2 = \Phi_{(XY)z}^2 \frac{1 - \Phi_{XY}^2}{1 + 2\Phi_{YX}\,\Phi_{X(XY)}^2 - 2\Phi_{X(XY)}^2 - \Phi_{XY}^2}$$

or

$$\Phi_T^2 = \frac{\Phi_{(XY)z}^2 (1 + \Phi_{XY})}{(1 + \Phi)_{XY} - 2\Phi_{X(XY)}^2} \tag{7}$$

Now the general formula for a product moment correlation coefficient is

$$\Phi_{ij} = \frac{C_{ij} - M_i M_j}{\sigma_i \sigma_j}$$

where C_{ij} is $\frac{\Sigma X_i X_j}{N}$. (8)

From 7 and 8

$$\Phi_T^2 = \frac{\Phi_{(XY)Z}^2 \left(1 + \frac{C_{XY} - M_X M_Y}{\sigma_X \sigma_Y} \right)}{\left(1 + \frac{C_{XY} - M_X M_Y}{\sigma_X \sigma_Y} \right) - 2 \left(\frac{C_{X(XY)} - M_X M_{XY}}{\sigma_{YX} \sigma_X} \right)^2} \quad (9)$$

But because of 4 and since all variables are dichotomous

$$\sigma_X = \sigma_Y = \sigma_Z = 1/2 \quad (10)$$

Also we can show that

$$C_{X(XY)} = C_{XY} = M_{XY} \quad (11)$$

Using 4, 10 and 11 in 9

$$\Phi_T^2 = \frac{\Phi_{(XY)Z}^2}{1 - \frac{C_{XY}}{2\sigma_{XY}^2}} \quad (12)$$

Or applying the general form 8 to 12.

$$\Phi_T^2 = \frac{2(C_{(XY)Z} - M_{XY} M_Z)^2}{\sigma_Z^2 (2\sigma_{XY}^2 - C_{XY})} \quad (13)$$

Now because X and Y are dichotomous it is easy to show that

$$\sigma_{XY}^2 = C_{XY} - C_{XY}^2 \quad (14)$$

Substituting 14 in 13 and using 4 and 10

$$\Phi_T^2 = \frac{4(2C_{(XY)Z} - C_{XY})^2}{C_{XY}(1 - 2C_{XY})} \quad (15)$$

Because of 4

$$C_{XY} = \frac{C_{X_1Y_1} + C_{X_2Y_2}}{2} \quad (16)$$

where $C_{X_1Y_1}$ and $C_{X_2Y_2}$ refer to the two subgroups respectively of the dichotomous criterion.

Also because of 4

$$C_{(XY)Z} = \frac{C_{(X_1Y_1)Z_1} + C_{(X_2Y_2)Z_2}}{2} \quad (17)$$

But since all Z_1 values are "1" and all Z_2 values are 0, we have $C_{(X_2Y_2)Z_2} = 0$ and 17 becomes

$$C_{(XY)Z} = \frac{C_{X_1Y_1}}{2} \quad (18)$$

Substituting 16 and 18 in 15

$$\Phi_T^2 = \frac{(C_{X_1Y_1} - C_{X_2Y_2})^2}{(C_{X_1Y_1} + C_{X_2Y_2}) \left[1 - (C_{X_1Y_1} + C_{X_2Y_2}) \right]} \quad (19)$$

Now assuming that each Z group was equally divided on its dichotomous responses to both the X and the Y variables, it can be shown that

$$\Phi_1 = \frac{C_{X_1Y_1} - 1/4}{1/4}$$

or

$$C_{X_1Y_1} = \frac{\Phi_1 + 1}{4} \tag{20}$$

where Φ_1 is the phi-coefficient between X and Y for the group with scores of 1 on Z. Similarily if we let Φ_2 be the correlation of those with scores of 0 on Z, then

$$C_{X_2Y_2} = \frac{\Phi_2 + 1}{4} \tag{21}$$

Substituting 20 and 21 in 19 gives

$$\Phi_T^2 = \frac{(\Phi_1 - \Phi_2)^2}{(2 + \Phi_1 + \Phi_2)[2 - (\Phi_1 + \Phi_2)]} \tag{22}$$

or

$$\Phi_T = \frac{\Phi_1 - \Phi_2}{\sqrt{4 - (\Phi_1 + \Phi_2)^2}} \tag{23}$$

which is the same as equation 1 in the text.

We have next to find the regression coefficients in the equation

$$t = B_0 + B_1X_1 + B_2X_2 + B_{12}X_1X_2 \tag{24}$$

Since B_0 is an additive constant and will not affect the correlation of t with the dichotomous criterion we shall consider first the solution for the other regression coefficients. Instead of the matrix and vector of correlations in 1 and 2, we consider the corresponding variances and covariances as in 25 and 26 respectively.

$$P = \begin{bmatrix} \sigma_X^2 & P_{XY} & P_{X(XY)} \\ P_{XY} & \sigma_Y^2 & P_{Y(XY)} \\ P_{X(XY)} & P_{Y(XY)} & \sigma_{XY}^2 \end{bmatrix} \tag{25}$$

$$P_z = \begin{bmatrix} P_{Xz} \\ P_{Yz} \\ P_{(XY)z} \end{bmatrix} \tag{26}$$

Also we let

$$B = \begin{bmatrix} B_1 \\ B_2 \\ B_{12} \end{bmatrix} \tag{27}$$

Then it is well known that

$$B = P^{-1}P_z \tag{28}$$

But because of 3, 5 and 10 it can be shown that in scalar notation

$$B_1 = P_{X(XY)}(P_{XY} - \sigma^2_X) \quad \frac{P_{(XY)z}}{D}$$

$$B_2 = P_{X(XY)}(P_{XY} - \sigma^2_X) \quad \frac{P_{(XY)z}}{D} \tag{29}$$

$$B_{12} = (\sigma^4_X - P^2_{XY}) \quad \frac{P_{(XY)Z}}{D}$$

where D is the determinant of P.

We let

$$K = (\sigma^2_X - P_{XY}) \quad \frac{P_{(XY)Z}}{D} \tag{30}$$

Substituting 30 in 29 gives

$$\begin{aligned} B_1 &= -P_{X(XY)}K \\ B_2 &= -P_{X(XY)}K \\ B_{12} &= (\sigma^2_X + P_{XY})K \end{aligned} \tag{31}$$

But by definition

$$\begin{aligned} P_{X(XY)} &= C_{X(XY)} - M_{XY}M_Y \\ P_{XY} &= C_{XY} - M_XM_Y \end{aligned} \tag{32}$$

Using 4, 10, 11 and 32 in 31

$$\begin{aligned} B_1 &= -C_{XY}(1 - M_X)K \\ B_2 &= -C_{XY}(1 - M_X)K \\ B_{12} &= C_{XY}K \end{aligned} \tag{33}$$

We let

$$G = \frac{K\,C_{XY}}{2} \tag{34}$$

Substituting 4 and 34 in 33

$$\begin{aligned} B_1 &= -G \\ B_2 &= -G \\ B_{12} &= 2G \end{aligned} \tag{35}$$

Substituting 35 in 24

$$t = B_0 - G(X_1 + X_2 - 2X_1X_2) \tag{36}$$

Now any linear function of t will correlate the same with Z as t itself. Hence we write

$$T = At + E \tag{37}$$

where

$$\begin{aligned} A &= 1/G \\ E &= 1 - B_0/G \end{aligned} \tag{38}$$

From 36, 37 and 38

$$T = 1 - X_1 - X_2 + 2X_1X_2 \tag{39}$$

which is the same as equation 5 in the text. We have therefore proved that the weighting of the variables X_1, X_2, and X_1X_2 indicated by 39 results in the maximum correlation with the criterion.

References

1. Gaier, Eugene L. and Lee, Marilyn C. Pattern Analysis. The Configural Approach to Predictive Measurement. *Psychol. Bull.* 1953, *50*, 141-149.
2. Meehl, P. E. Configural Scoring. *J. consult. Psychol.*, 1950, *14*, 165-171.
3. Stouffer, Samuel A., Borgatta, Edgar F., Hays, David G. and Henry, Andrew F. A Technique for Improving Cunulative Scales. *Public Opinion Quarterly*, 1952, *16*, No. 2. Summer, 1952.

A FACTOR ANALYSIS OF CURVILINEAR DISTORTIONS ON THE BENDER-GESTALT[1]

WILSON H. GUERTIN

Veterans Administration Hospital, Knoxville, Iowa

INTRODUCTION

The Bender Visual Motor Gestalt Test[1] which was introduced in 1939 has found ready application with mental hospital patients. Interpretations are largely intuitive and based upon clinical experience with the instrument. Billingslea[2], Hutt[5], and Pascal and Suttell[6] have worked toward more unified scoring and interpretive schema. In a previous paper by the author[4] an attempt also was made to see if some systematic order could be developed out of the relationships among scoring items through the application of factor analysis. Five fairly meaningful factors were extracted at that time from a rather sparse sampling of scoring categories.

One cluster of scoring categories embodied curvilinear distortion and rotational tendencies. Since the corresponding factor was most heavily loaded with mixed and catatonic schizophrenics, the hypothesis was suggested that poor emotional control underlies certain types of distortions. Furthermore, since the cluster of variables seemed heterogeneous, it was expected to be subject to further simplification by selecting pertinent variables and subjects. The present report is of a study attempting to include all the scoring categories that might be particularly related to this cluster.

PROCEDURE

The sample consisted of 100 white mental patients at several hospitals, 26 females and 74 males. These subjects were classified grossly for statistical convenience into five diagnostic classifications. The distribution of subjects by diagnostic groups was: 13 non-schizophrenics, 23 paranoid schizophrenics, 15 catatonic schizophrenics, 14 hebephrenic schizophrenics, and 35 mixed schizophrenics. The non-schizophrenic group included character disorders and organic brain-damage cases in about equal proportions.

A total of 46 scoring variables were studied, four of which had to be rejected because of an inadequate number of cases in the tails of the distributions. Nine of the variables retained showed tails with frequencies of occurrence of 11 to 20 percent of the cases. The scoring items employed stem from distortions encountered rather frequently in the productions of hospitalized schizophrenics. Thus, it can be seen that the study is directed toward an understanding of the factors underlying the productions of schizophrenics rather than a more generalized group of patients. Furthermore, only those deviations were scored that were sufficiently gross distortions to be recognized without measuring techniques.

The information from the data sheets on each subject were transferred to McBee Punch Cards for ease of tabulation. Tetrachoric correlation coefficients were graphically estimated from Thurstone's Tables. The multiple group centroid method of factor analysis[7] was employed because it is time-saving and likely to produce empirically meaningful clusters[3]. The clusters which determined these multiple group factors were identified by a conventional method which produces fairly tight and relatively independent clusters. Communality estimates were taken as the highest coefficient in each column except for those tests in a cluster. The communality estimates for cluster tests were taken as the highest intercorrelations between the particular tests and any other member in the same cluster. The communality accounted for by each of the variables was arrived at from an orthogonal matrix which was obtained by rotating the oblique factor matrix.

[1]From the V. A. Hospital, Knoxville, Iowa. The author is indebted to Dr. M. C. Addington of the Clinical Psychology Section, V. A. Medical Teaching Group, Memphis, for providing a portion of the tests herein employed.

Results

While extremely high correlations were absent from the intercorrelation matrix, six clusters were quite easily identified and appeared to be relatively independent of one another. The six clusters determined the oblique matrix with the six factors abstracted simultaneously. This resulting six-factor oblique matrix appears as Table 1. Intercorrelations between factors were quite low. The highest were −.32 between A and C, .32 between A and D, and −.30 between B and E (factors are lettered in their order of appearance in Table 1).

Table 1. Oblique Factor Matrix*

Variable	Figure	Factors						
		Unstable Cloture	Curvilinear Distortion	Propensity for Curvilinear Movement II	Fragmentation	Irregular Prop. for Curv. Mvmt.	Experimental Dependence	Σa_{ij}**
Lips not approximated	5	87	−03	−23	04	−14	−04	85
Lips disapproximated	5	76	01	−42	34	−11	−13	71
Cup asymmetric	5	68	16	−09	37	−05	08	52
Cup lengthened	5	−45	−17	18	−08	09	−07	23
Angles rounded	3	28	−05	−15	−18	18	−22	24
Square rotated	A	−22	14	−01	−01	−01	−16	11
Cup squared	5	21	−09	09	−04	−06	12	11
Major arc asymmetric	4	19	−12	−02	10	05	−01	09
Angle between elem. wrong	6	17	10	01	−01	06	−02	07
Contour asymmetry	6	−06	87	−14	−04	−34	07	82
Humps angled	6	−02	85	12	26	−29	11	84
Circle asymmetric	A	36	55	−20	21	−06	11	42
Short major arc	4	35	36	−15	33	23	−03	38
Condensed dots, 1 side	5	−05	35	01	−02	−14	−27	26
Square rotated	4	16	28	08	08	−27	24	19
Minor arcs self-prop.	4	−49	−12	79	−10	01	−01	74
Diamond rotated	8	02	−09	69	−01	−39	31	65
Horizontal elem. rotated	6	−21	−32	65	−12	−10	−15	48
Too many humps	6	−02	09	56	−02	−47	−10	39
Non-closure	A	22	36	−16	73	10	04	58
Handle rotation	5	13	−11	−01	59	−03	−08	47

TABLE 1. OBLIQUE FACTOR MATRIX*—(*Continued*)

Variable	Figure	Factors						
		Unstable Cloture	Curvilinear Distortion	Propensity for Curvilinear Movement II	Fragmentation	Irregular Prop. for Curv. Mvmt.	Experimental Dependence	$\Sigma aij2$**
Non-closure	4	28	18	-04	58	11	03	37
Attachment point wrong	A	-18	22	-07	49	26	19	49
Overlap	4	-14	09	-02	-49	37	20	66
Extension rotated	5	-09	16	01	47	19	-06	33
Overlap	A	-43	25	-01	-44	-17	-04	44
Humps flat	6	03	30	-21	-40	-05	02	38
Angles asymmetric	3	-01	-05	-15	23	-04	15	17
Vertical element rotated	7	02	08	01	19	11	-09	06
Humps too few	6	06	-18	-33	11	72	-15	55
Horizontal elem. rotated	8	-10	-20	-08	06	62	-05	40
Angle between elem. wrong	7	06	-26	-03	06	60	-18	38
Circle unclosed	A	-12	02	-22	11	41	-30	28
Crossing point wrong	6	27	39	-12	22	-38	15	31
Circle smaller	A	-01	11	00	21	-22	-04	13
Cup rotated	5	-03	04	-08	21	18	11	10
Curve not centered	4	-10	24	11	06	17	03	20
Minor arcs not small	4	-27	-08	16	-26	-22	81	78
Minor arcs large	4	02	22	08	08	03	81	77
Major arc small	4	16	14	09	18	-27	59	44
Major arc large	4	15	06	03	04	08	-26	13
Curve rotated	4	-03	25	00	22	-09	-26	20
								$\Sigma aij^2 = 16.72$

*Decimal points omitted in all tables.

**Based upon orthogonal matrix not reported here.

The total variance accounted for by the six orthogonal factors constitutes 40 percent of the total variance of the intercorrelation matrix or 78 percent of the estimated communality for these 42 variables. The estimated communality for the intercorrelation matrix was 51 per cent of the total variance.

TABLE 2. FACTOR LOADINGS FOR DIAGNOSTIC GROUPS

Diagnosis	Factors					
	Unstable Cloture	Curvilinear Distortion	Propensity for Curvilinear Mvmt. II	Fragmentation	Irreg. Prop. for Curvil. Mvmt.	Experimental Dependence
Paranoid Schizophrenia	.00	−.48	−.03	−.52	.17	−.28
Catatonic Schizophrenia	.39	.06	−.16	.14	.06	.07
Hebephrenic Schizophrenia	−.28	−.09	.35	.04	−.40	.04
Mixed Schizophrenia	−.07	.33	−.01	.49	−.04	.43
Other	−.19	.38	.05	−.16	−.17	−.03

Loadings on these derived factors for diagnostic groups are shown in Table 2. It will be noted that there are a fair number of moderately high loadings. Some assistance in understanding the derived factors can be sought in their relationships to the previously reported Bender-Gestalt factors [4]. Mean correlations for cluster items of the previous study with the clusters in this study led to the correlations given between them in Table 3.

TABLE 3. INTERRELATIONSHIPS BETWEEN OBTAINED FACTORS AND THOSE FROM PREVIOUS STUDY

Factors from Previous Study	Factors					
	Unstable Cloture	Curvilinear Distortion	Propensity for Curvilinear Mvmt. II	Fragmentation	Irreg. Propen. for Curvil. Mvmt.	Experimental Dependence
Poor Reality Contact	.05	.09	.15	.25	−.17	−.09
Careless Execution	.07	.46	−.18	.31	−.19	−.02
Constriction	−.01	.20	−.07	.08	−.08	−.03
Poor Spatial Contiguity	−.10	−.02	−.15	.13	.46	−.03
Propensity for Curv. Mvmt. I	.04	.19	−.04	.30	.30	.06

DISCUSSION

This study, like the previous one, suggests that Bender-Gestalt performance can be understood fairly well in terms of common factors, since the communality of the intercorrelation matrix is slightly greater than half the total variance of the matrix. Such findings suggest that it would be profitable to employ clusters (or factors) as scoring categories, instead of using individual distortion items, when analyzing the performance of mental patients on the Bender-Gestalt.

Rotations were not attempted inasmuch as the conventional cluster analysis employed would lead to a first approximation to simple structure. The apparent

meaningfulness of those clusters is further evidence that the positions of the reference axes are satisfactory.

The six factors have been named inductively by considering the symptoms which showed the highest factor loadings.

The Unstable Cloture Factor is confined almost entirely to Bender figure 5. One might suspect that this factor is largely an experimental dependence item, but the moderately high loading of .39 for catatonics on this factor suggests that it may be something more. It is characterized by a disapproximation of the lips of the cup in figure 5, which tends to reduce the likelihood of making the cup symmetric because of the larger distance spanning between the edges of the cup. Just why an individual would avoid approximating the edges of the cup and thus produce an unstable and asymmetric figure is not clear. It would appear that some emotional elements operate to produce such an unstable configuration. The catatonics are the only ones positively loaded with this factor and it seems likely that this factor is related to an underlying general emotional instability. The highest negative loading is among the hebephrenics, and while they present emotional instability, it is more in the nature of a directly observable situational response rather than a general disturbance of emotional balance.

The Curvilinear Distortion Factor is characterized by executional inaccuracy of curvilinear drawing. It may well be that the execution of curvilinear drawing has some sensual qualities which provoke impulsive responses. Thus, it is proposed tentatively that these curvilinear distortions are caused by impulsiveness with a possible emotional basis. This factor corresponds rather closely to a factor from the previous study which was concerned with the careless execution of more general aspects of the Bender-Gestalt. Further study will be required to see whether the two factors can be differentiated from each other.

The Propensity for Curvilinear Movement Factor II seems to be an essentially psychotic factor. Not only are rotations involved, but a generalized propensity for curvilinear movement seems to operate in the drawing of the two minor arcs on figure 4. These are well executed by individuals possessed of this factor inasmuch as they find a certain freedom in curvilinear expression with inherent change of direction. This facilitates the reproduction of each arc so that they are proportionate to one another. The emotional disorganization and affect display which seem to underly this factor are reflected in the loading of the hebephrenics on this factor.

The Fragmentation Factor also appears to be an essentially psychotic factor. It is characterized by a disjoining of the elements of the figure and either represents a misperception of the elements as being separated or a primitive defense to avoid feeling through dissociation. This factor seems to be primarily a factor concerned with defenses against feeling, perhaps largely against anxiety rather than emotions. Psychiatrically, this type of defense would be characterized as dissociation, except in the chronic deteriorated schizophrenic for whom the term dissolution would be more apt.

The Irregular Propensity for Curvilinear Movement Factor seems to be a higher form of defense against feeling. We find that it is not unlike the Poor Spatial Contiguity Factor found in the previous study, which was encountered in character disorders and neurotics. These individuals seem to be experiencing conflict over emotions which are unacceptable to the ego. They frequently respond with neurotic-like mechanisms against these feelings, such as reaction formations. Despite the description of these as "neurotic-like defenses" it is seen in Table 2 that the mixture of organics and character disorders ("non-schizophrenics") do not utilize these particularly, with the highest positive loading among the paranoids. Hebephrenics do not have the intellectual capabilities requisite for employing such high-level defense mechanisms and, as expected, are highly loaded negatively on this factor.

The Experimental Dependence Factor seems to stem entirely from inter-relations of scoring items on figure 4. Since the arcs were judged with respect to one another as to whether they were disproportionately large or small, the largest portion of the common variance is attributable to the scoring and the mutually exclusive possibilities inherent therein. However, a rather specific factor related to the disproportion between the small major arc and the large minor ones seems to have some importance for the group of mixed schizophrenics. Its meaning is most difficult to evaluate from this study alone. In general, it would be wise to disregard this last factor until it may reappear in a future study.

In summary, it may be stated that this attempt to factor analyze variables related to curvilinear movement has disclosed at least some factors different from those obtained in a more gross analysis. There is reason to suspect that these obtained factors hold more diagnostic significance for the evaluation of emotions than do the factors previously reported [4]. The psychological meaning of these factors, unfortunately, is less clear than in the previous study. Only further cross validation with internal checks for psychiatric symptomatology can provide sound principles of interpretation for the Bender-Gestalt productions of mental patients. Those psychological significances attached to the factors in this paper are highly tentative, and while suggestive, should not be applied carelessly in clinical situations.

Summary and Conclusions

1. The previous factor analysis of the Bender-Gestalt [4] produced a *Propensity for Curvilinear Movement Factor I* which seemed to be related to emotions. The present study was directed toward a further evaluation of propensity for curvilinear movement through a more restricted sampling of scoring items and patients.

2. The Bender productions of 100 mental hospital patients were evaluated in terms of a number of semi-objective scoring indices. The intercorrelation matrix of these variables was factored with the multiple-group centroid factor analytic technique. Factor loadings are presented for 42 test variables and five gross diagnostic categories. Relationships between the factors obtained from the current study and a previous one are also presented.

3. The five oblique factors obtained which seem to have meaning are: the *Unstable Cloture Factor* related to general emotional balance, the *Curvilinear Distortion Factor* related to impulsiveness, the *Propensity for Curvilinear Movement Factor II* related to emotional disorganization and affect display, the *Fragmentation Factor* related to dissociation or disorganization, and the *Irregular Propensity for Curvilinear Movement Factor* which is related to emotional conflicts and neurotic-type defenses against feelings.

4. The results of the study were regarded as encouraging, and it is felt that an integrating study incorporating behavioral data may lead to a much sounder diagnostic approach to the Bender-Gestalt productions of mental hospital patients.

Bibliography

1. Bender, Lauretta. A visual motor-gestalt test and its clinical use. *Amer. Orthopsychiat. Assoc., Res. Mongr.*, 1948, No. *1*.
2. Billingslea, F. Y. The Bender-Gestalt: An objective scoring method and validating data. *Clin. Psychol. Monogr.*, 1948, No. *1*.
3. Cattell, R. B. *Factor Analysis.* New York: Harper & Bros., 1952.
4. Guertin, W. H. A factor analysis of the Bender-Gestalt tests of mental patients. *J. clin. Psychol.*, 1952, *8*, 363-367.
5. Hutt, M. L. *A tentative guide for the administration and interpretation of the Bender-Gestalt test.* U. S. Army: Adjutant General's School, 1945 (Restricted).
6. Pascal, G. R., and Suttell, B. J. *The Bender-Gestalt test quantification and validity for adults.* New York: Grune & Stratton, 1951.
7. Thurstone, L. L. *Multiple-factor analysis.* Chicago: University of Chicago Press, 1947.

A STORY COMPLETION TEST FOR COLLEGE STUDENTS

EUGENE S. MILLS

Whittier College

In an earlier article [4] the writer discussed a story completion test which was designed for individual use with children. The originator of this test [7] found the story completion technique to be profitable clinically and this has been supported by subsequent application to children of different ages.

In view of the fact that psychologists must deal with large numbers of college students and that they often must obtain material which will assist in making evaluations of student personality, it would appear worthwhile to attempt to adapt the story completion technique to the age level and life experiences of the college student. It would seem particularly promising if the technique could be shown to work effectively in group testing.

In order to try out the basic story completion technique in group testing at the college level, the writer composed over fifty story completion items dealing with various critical areas of student life. In deciding upon the areas of student life to be represented in the items, a study was made of several research projects which dealt extensively with the life and problems of college students [1, 2, 3, 5, 8]. The seven general areas under which the items might be subsumed were: (a) home, (b) academic, (c) social, (d) vocational, (e) religion and philosophy of life, (f) athletic, and (g) health. The structure of the items was similar to that of the Madeleine Thomas Completion Stories Test items.

Different combinations of items were administered to psychology students in two liberal arts colleges. An attempt was then made to select those items which analysis showed to be most productive. Many items were eliminated because of the perfunctory or stereotyped responses which they seemed to elicit. In general, however, it was found that the students took the items seriously and responded in an open and spontaneous manner as long as the test was introduced with the proper guarantees as to the scientific and confidential nature of the material.

After having experimented with each item, a test of fifteen items was established closely paralleling the Madeleine Thomas Test. The criteria for inclusion of an item were: (a) it must be maximally productive, and (b) it must complement the other items in the test. The test items included produced, as a unit, a wealth of material when administered to a final group of history and psychology students.

The Test and its Administration

Following are the fifteen items which the writer has found to be most productive in group testing of college students:

1. A student is in college. He[1] writes a letter home. What does he say in the letter?
2. This student answers a question incorrectly in class. In the hall after class another student says to him: "You sure messed that one up, didn't you?" What does he do?
3. The student calls at the college post office where he finds a letter waiting for him. It is from his parents. What does it say?
4. The student has been to a movie with his best girl friend. After the date he returns to his room in the dorm. How does he feel? What does he do?
5. The student has been going steady with a certain girl. One night she refuses to go out with him. Why? What does he do?
6. The student tries out for the football team[2]. What happens? How does he feel?

[1] In use with women students it is helpful to make proper substitutions for the sex of the hypothetical student in the stories.

[2] In use with women students the substitution of the term "volley ball team" facilitates the response to this item.

7. The student receives his final grades for the semester. How does he feel? What does he do?
8. The student has been sick in the college infirmary for a few days. After giving him an examination, the college physician says . . . . ?
9. It is night, the student is in bed, the day is over. What is he thinking about?
10. Then he goes to sleep. What does he dream about?
11. The student is now a senior. Is he eager to graduate or would he rather remain in college for a while? Why?
12. What will he do after he graduates?
13. Among all the novels and stories he has heard and read, which one does he like the best of all? Why?
14. If you were to write a book, what would you write about?
15. Do you remember your early childhood? What is the first thing you can remember now?

In preparing the test for use with groups of students it has proven helpful to spread the items over several pages, allowing ample space between each item (and between each section of the items) for the student to write his answers. It has also seemed necessary to urge the student to extend his answers to the reverse side of the pages when necessary, after having indicated the number of the item he is answering.

In using the story completion test with groups of students much of the value of the test depends upon the extent to which the students have been prepared for the test items. It seems desirable to talk informally with them prior to the test about the nature of the task, giving them assurances of the confidential nature of the material, and urging them to enter freely into the stories. In practice, the writer has made introductory comments such as the following:

> You are going to be given some pages on which are written a number of short stories or items about a college student. Each story describes a situation and asks a question about that situation. You are to read the story item and then complete it in any way you wish, extending your completion of the story to the back of the page if space allowed is inadequate. When doing so, however, please indicate the number of the story you are completing. Please feel free to use your imagination in completing the stories because there are, of course, no right or wrong endings for them.

Method of Analysis of Test Responses

The technique which has been used for analysis of test responses is essentially the same as that devised by the writer [4] for analysis of the Madeleine Thomas Test and is similar to that used by Symonds [6] in his research on adolescent fantasy. In essence, the task was one of theme analysis of all story completions. The writer's experience has been the same as Symonds who found that, "No 'system' seemed to be wholly satisfactory. Nothing took the place of giving the task one's individual attention, keeping as many details as possible in memory, and maintaining a flexible attitude" [6, p. 71].

In analyzing the fantasy material produced in response to the stories, the writer removed all notations which served to identify the subject who had produced those responses in order to minimize the possibility of prejudice in his evaluation. A general survey was made of all responses in order to become acquainted with the subject's style of writing and his orientation toward the test stories. Following this, each story completion was read slowly, particular attention being paid to recurring themes, omissions, confused writing, witty comments, and other indications of disturbance. Recurring themes and other qualitative observations were noted in the margins next to the story or stories concerned.

After each response had been thus studied, a further general survey of the test as a whole was made in an attempt to relate notations to the over-all body of the test. Following this operation, the responses were reread in reverse order and further notations made as necessary. Two days later the writer came back to the test material and repeated the above operations. An attempt was made to distill further the qualitative comments which had been made previously, sharpening the nota-

tions to describe more accurately the dynamic qualities of the responses and eliminating unnecessary duplication in noting themes.

It must be stated, of course, that the writer has not always been certain as to how to characterize specific themes. Often the projective material is such as to defy classification, while at times it is difficult to decide whether or not material should be listed as two separate themes or as one theme. While Symonds[6] chose to list themes only when they appeared in the projective material at least three times, the writer has made no such limitation in recording theme material. In effect, it often appears that a particular theme occurs just once in a subject's test, but that in light of other test indications such a theme constitutes a significant element in his adjustment.

Results

At the present time the test has been administered to 312 history and psychology students in three liberal arts colleges and the responses analyzed in accordance with the general technique just discussed. A total of 6,864 themes have been tabulated according to the number of the story item where they appeared. Table 1 presents the percentages of themes of different types which appeared in completion of the separate items.

Table 1. Percentages for Theme Analysis of Responses to Each Story Item (Figures Rounded to the Nearest 1%)

	Story Items														
Themes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Family relationships	52		52						3	7			5	17	36
Student relationships	15	7	3	20	2	6	1		7	6					
Academic problems	6	40	15	8	1	7	49	7	10	12					
Level of aspiration		5	3	2		18	4		3		12	10	17	13	
Financial status	6		7	1		2	3		2	2	8	17	11	10	6
Guilt	3	17		8	16		12		3	7					10
Health	8		3		3	14	2	67	12	2	6				12
Sex				42	27			8	9	10		1	17	14	
Fears	2	5	3			6	11	15	2	2	20			6	21
Vocation	1		1				2		3	3	37	39	2	9	
Fantasy			2	5	3	1			10	29		11	31	20	
Rejection	1	6	4		37	23	9		6	1	2	7			7
Aggression		17	1	6	6	10		2	5	2		2	6		2
Religion and philosophy of life				3					10	11	5	7	7	8	
Traumatic	1	3	1	1	4	6	1		1	1	1		1		1
Homosexuality									3	1					
Miscellaneous	5		5	4	1	7	6	1	11	4	9	6	3	3	5
Total	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

It is apparent from this table that a variety of themes was present in the responses to the test and that certain items are more likely to elicit a particular theme (or themes) than are other items. Examples of this are seen in the case of items 1 and 3, in which approximately one-half of all themes concern family relationships, and items 11 and 12, which produce a wealth of material concerning vocational matters. While both items 1 and 3 elicit material dealing with family relationships, the different approach in the two items makes possible an instructive comparison of the student's attitudes with what he feels to be the attitudes of his parents. Since family relationships are of great importance in the lives of most college students, this test condition facilitates an understanding of many student problems. Similar inter-item comparisons are possible in other critical areas of adjustment.

There is, of course, considerable difference in the time required for students to complete the story items. On 246 tests which were timed to the nearest minute, the range was 7 to 59 minutes. It has been observed that the more seriously disturbed students tend either to complete the test in a relatively short time or, as is sometimes the case, to spend an excessively long time on the task.

A limited test of this observation was possible with the data at hand. Rorschach records to which the Munroe Inspection Technique had been applied were available on 43 cases. On the basis of the size of Munroe scores, the writer ranked the records from the most seriously maladjusted to the most favorably adjusted. By arbitrarily selecting the top ten, which Munroe scores revealed to be the most poorly adjusted, and plotting these ten cases against the time taken to complete the story completion test, it becomes apparent that they tend toward the extremes. Of the ten cases studied, six took fifteen minutes and under to finish the test, while the remaining four took forty-five to fifty-nine minutes. The time required by these ten cases stands in rather striking contrast to the mean time of 32 minutes for the 246 cases tested. This suggests that the time factor may, in itself, be a significant clue in the use of such a technique.

Illustrative Case

The test has been used primarily with groups of college students; however, it has also been helpful to the writer in individual use with students who are experiencing difficulty in adjusting to college life. Once rapport has been established with the disturbed student the administration of the story completion test by the counselor in a face to face relationship offers a convenient entry into the student's problems. The experience of providing stories for the items often facilitates the recognition and discussion of critical areas of adjustment. An example of this is found in the case of a 20 year old student who came to the writer complaining of a continuing inability to decide upon a vocation. Following a brief period of introductory discussion the story completion test was administered orally to the student. Answers to three of the items illustrate the useful material obtained by this method.

Item 1. (A student is in college. He writes a letter home. What does he say in the letter?) "I guess he says, Dear Folks: I'm getting along O. K. Uh . . . my studies are O. K., too, but I can't seem to get an idea of what to do with them. Here I am—I'm in college and I don't know what I want to do yet. I like my science courses but I make better grades in English and Drama. Please don't lose patience with me . . . I'll try to get situated soon. I could use some more dough. Love, Johnnie."

Item 3. (The student calls at the college post office where he finds a letter waiting for him . . . It is from his parents. What does it say?) "Dear Johnnie: Er . . . I got your note. Sorry you haven't had any luck getting your mind made up yet. We're paying for your . . . we're paying your way and naturally we'd like to see some progress on this. Your mother and I hope you will keep in mind . . . uh . . . our talk at Christmastime. Love, Dad."

Item 11. (The student is now a senior. Is he eager to graduate or would he rather remain in college awhile? Why?) "He'd like to, I guess, get out . . . but he also doesn't look forward to the big push after the protection of college walls. You know . . . ivory towers and all that stuff. He's got to make up his mind soon . . . or someone else will make it up for him."

In this case the story completion responses provided much insight into the difficulties besetting the student. It was possible to infer from the responses that the parents were significantly involved in the case. The talk mentioned in response to item three was a clue which was followed up in a subsequent interview. In this case the parents had been pressuring their son to study law in the expectation that he would some day join his father's law firm. During the previous Christmas vacation the father had promised his son a new car if he would apply for admission to law school. The degree of anxiety and concern which the student felt was amply demonstrated in the story completion responses along with valuable clues as to prevailing conditions in his case.

The story completion test does not appear highly threatening to most students. For this reason it has been useful as a first test in those cases where considerable testing seems indicated. Besides providing valuable information early in the relationship, it helps to clear the way for the Rorschach, T. A. T., and other tests which usually constitute a greater threat.

Summary

The writer discussed a story completion test for college students consisting of 15 items dealing with critical areas of student life. The test was administered to 312 history and psychology students in three liberal arts colleges and a theme analysis made of the responses. The time required for administration of the test was discussed and consideration given to the value of the technique as an opening wedge in testing and counseling relationships.

References

1. Fry, C. C. *Mental health in college.* New York: The Commonwealth Fund, 1942.
2. Katz, D. and Allport, F. H. *Students' attitudes—a report of the Syracuse University reaction study.* Syracuse, N. Y.: The Craftsman Press, 1931.
3. Lloyd-Jones, E. *Social competence and college students.* Washington, D. C.: American Council on Education, 1940.
4. Mills, E. S. The Madeleine Thomas Completion Stories Test. *J. consult. Psychol.*, 1953, 2, 139-141.
5. Munroe, Ruth. *Adjustment and academic performance of college students.* Stanford University: Stanford University Press, 1945.
6. Symonds, P. M. *Adolescent fantasy.* New York: Appleton-Century-Crofts, 1951.
7. Thomas, Madeleine. Methode des histoires a completer. *Arch. de Psychol.*, 1937, 26, 209-284.
8. Wood, B. D. and Learned, W. S. *The student and his knowledge.* New York: The Carnegie Foundation for the Advancement of Teaching, 1938.

PERSONALITY INTEGRATION AND THE PERCEPTUAL PROCESS[1]

ELI A. RUBINSTEIN

Veterans Administration, Washington, D. C.

Problem

This investigation concerns the influence of personality integration on the perceptual process by examining the relationship between personality integration, as a quantitative aspect of personality dynamics, and the perceptual process. Is the perceptual process increasingly influenced by internal, personal factors of the perceiver as personality integration decreases?

Previous studies in this field relate to this investigation more by implication than directly [1, 2, 3, 5, 7]. In their sum total they lend support to the hypothesis that degree of personality adjustment influences perception, since the greater the personality maladjustment the more probably there will be strong somatic needs, intense personal values and attitudes, or strong emotional tensions to influence the percept.

We can hypothesize two broad continua entering into the determination of relative weights of internal and external factors in the perceptual process. On one hand, the degree of structuration of stimulus should largely underly the influence of the external stimulus. On the other hand, the degree of personality integration should largely underly the pressure of the internal factors. Given a series of stimuli of graded ambiguity, the greater the structuredness, the greater the conformity in response. But, at the same time, given a group of perceivers, the greater the personality maladjustment, the less the conformity in response (assuming the maladjustments are not all identical in nature.) If these hypotheses are valid, we should be able roughly to predict the result of presenting a series of stimuli of graded ambiguity to a group of individuals of varying personality adjustment. We would expect that the greater the structuredness of stimulus, the greater the conformity of response. But, in addition, we should expect to find, as a modifying effect in this framework, that the greater the personality disturbance, the less the conformity of response. The better the personality adjustment (and, therefore, the less the influence of individual internal factors), the earlier in the presentation of the series of graded stimuli will we receive the "expected" or conforming response.

Experimental Design

Stimulus Material. A series of incomplete line drawings of simple objects was devised and three graded stages of ambiguity were produced by successively stripping the incomplete drawings of key parts until only the barest implication of the drawing was left. Thus, in a sense, the series was developed backwards. As actually presented to the subject, each individual line drawing was on a card, 36 cards in all, and each card was shown to the subject for as long as needed to elicit a response. The experimenter had, at the same time, a recording sheet on which were reproduced the line drawings. He roughly sketched the response on the sheet in accord with the explanation of the subject, who was asked to trace his response, with his finger, on the card. Thus, no actual drawing was done by the subject. All first stage presentations were shown, then all second stage and then third stage presentations, rather than running through the three stages for each object successively.

The content of this graded-in-structuration series of line drawings was considered from two aspects. The attempt was, first of all, to have objects which, in complete form, have fairly definite and well-known configurations. Objects which lent themselves well to simple line representation were considered. In addition, the range of content was restricted to everyday, ordinary objects so as not to complicate

[1]This article is based partially on a doctoral dissertation submitted to The Catholic University of America. Grateful acknowledgments are due to Dr. William D. Commins for his assistance in the original study, and to Dr. A. C. Cornsweet for his cooperation in the continuation of the research.

the problem by the influence of cultural differences. As finally adopted, the set of stimuli included; a shoe, a hand, a profile, a fish, a dog, a butterfly, a watch, a tree, and a house. Thus, there were three human details, three animals, and three common objects.

It should be noted that this totals only nine presentations whereas the series actually contained 36 stimuli or 12 to each of the three stages. The additional three presentations per stage consisted of line drawings which became successively more complex, but at the same time less, rather than more, structured or unambiguous. Whereas the first nine drawings increased in complexity per stage toward a specific representational configuration, these three "nonsense" presentations increased in complexity away from any specific, well-known representation rather than towards such design. These three nonsense stimuli were introduced to hinder the establishment of a mental set in which the subject would, upon presentation of the second stage drawings, soon realize the increasing structuration and possibly withhold doubtful responses in anticipation of the even more structured third stage. It was also in order to prevent this mental set that the cards were presented as they were, rather than running through the three stages for each object successively. Furthermore, the cards were changed in sequence from stage to stage.

Subjects. In order to eliminate various cultural and environmental differences as much as possible, all the subjects were white male war veterans. On the possibility that organic brain damage might influence the test results, all brain damage cases, cases having undergone shock treatments of various kinds, or cases with any central nervous system disturbance were excluded from the study. Also, the Wechsler-Bellevue Intelligence Test was given to all subjects so that the influence of intellectual factors could be determined and statistically handled.

The sample consisted of 125 cases distributed over a fairly wide range of degree of personality adjustment. The subjects included patients of two mental hospitals, two mental hygiene clinics, and veterans applying to a university counseling center.

Criteria of Personality Integration. It was decided to use a single, over-all rating scale for the purpose of getting a measure of personality integration. A brief statement was constructed establishing the frame of reference in which the rating was to be made. Using that as a basis, a single five-step rating scale was devised as follows:

1. No indication of sufficient conflict, maladjustment, or personality disintegration to cause the patient to be considered abnormal.
2. A mild degree of conflict, maladjustment, or personality disintegration which prevents the patient from functioning on a completely adequate level.
3. A moderate degree of conflict, maladjustment, or personality disintegration which significantly restricts and makes unsatisfactory the patient's relationship with himself and his normal environment and which strongly suggests the need for some form of readjustive process.
4. A marked degree of conflict, maladjustment, or personality disintegration which makes it impossible for the patient to effect any adequate relationship at the present time with his normal environment without some comprehensive readjustive process.
5. An extensively severe degree of conflict, maladjustment, or personality disintegration which necessitates institutionalization or special therapeutic and custodial care, since even minimal socially adequate functioning is not possible on an independent basis.

This rating scale was filled out three times for each subject. In each case where a psychiatrist had extensive personal contact with the individual (this was the case for 66 individuals) the psychiatrist was asked to rate the individual. The psychologist who tested the person and wrote a report on him was asked to fill out another rating. The author, independent of the other ratings, reviewed the case history or social history of each individual and rated him a third time on the five point scale, taking into account such factors as: nature of illness, if any, length of illness, length and number of hospitalizations, nature of onset, description of behavior, and any other pertinent data presented. In addition to these ratings, each individual was given a "Basic Rorschach Score" based on the Buhler-Lefever[4] check list.

Results

Rating of Personality Adjustment. The product-moment correlation between the combined ratings and the Basic Rorschach Scores was .56, and showed that both do, to a considerable degree, measure the same thing. It was then decided to combine all four measures, the three ratings and the Rorschach Score, so as to have a combined rating for each individual. This was done by converting the Basic Rorschach Score distribution into five equal percentage groups to agree with the five-point rating scale and pooling all four scores into one total score.

The reliability was determined by average intercorrelation of all ratings as shown in Table 1. Guilford's Tables[6, p. 616] were used for converting correlations to Fisher's Z, averaging the Z's and then reconverting to a correlation coefficient. The average intercorrelation for all six correlations (three estimated uncurtailed distributions for the psychiatric ratings) was .75. This was one estimate of the reliability coefficient for the four sets of ratings.

Table 1. Intercorrelations Among the Four Ratings

	Psychiatrist		Psychologist		History	
	66 Cases	125 Cases	66 Cases	125 Cases	66 Cases	125 Cases
Psychologist	.76	.88*				
History	.64	.81*	.73	.92		
Rorschach	.38	.57*	.33	.52	.28	.50

*Estimated uncurtailed distribution (McNemar, p. 126)

The Graded Stimulus Test Scores. If the analysis and scoring of the graded stimulus test was based on a priori theoretical grounds, rather than empirical item analysis, and then correlated with the ratings afterwards, whatever measure of relationship would appear more probably would exclude the possibility of a positive bias due to unique factors in this particular sample. Scoring on theoretical grounds would also give a degree of independent meaning to the tests not obtainable by the above purely empirical method of scoring. For these reasons, the approach to the test analysis was made independent of the ratings.

As already explained, the stimuli used, except for three nonsense figures, were actual incomplete line drawings in three stages of ambiguity. The obvious first task was to tabulate the number of "correct" and "incorrect" responses. It should be expected over the entire sample that, despite the differences expected among the five ratings, there should be a definite and a sizable decrease in incorrect responses from Stage I, with greatest ambiguity, through to State III, with least ambiguity. Table 2 illustrates that such was the case. For the nine meaningful figures in the series it can be seen there was a definite and marked decrease in other than expected responses as structuredness increased from stage to stage. (The three nonsense stimuli, with no correct response, were treated somewhat differently, as will be explained.)

Even on purely theoretical grounds it would have been apparent that an incorrect response to a stimulus at Stage III should be weighted more heavily than an incorrect response to Stage I, when the stimulus is so much more ambiguous. It had been intended to weight the items, accordingly, with weights of 1, 2, and 3. However, in the final weighting, acknowledgement was made to the empirical evidence by giving relatively greater weight to the more infrequently found incorrect responses in Stage III. The weights assigned were 1, 2, and 5, for each of the three stages respectively.

The next area of significance was the series of three nonsense stimuli. On such perceptual tests as the Rorschach, rejections are considered important. This was a particularly appropriate approach in regard to the nonsense stimuli since they were

TABLE 2. "INCORRECT" RESPONSES BY ITEM AND BY STAGE

	Stage		
Item	I	II	III
Shoe	100	32	1
Fish	122	44	11
House	93	53	25
Animal	88	33	1
Tree	29	26	7
Profile	36	12	0
Watch	124	75	4
Butterfly	125	66	7
Hand	117	52	6
Totals	834	393	62

specifically designed to elicit no definite response. Accordingly, the scoring on these items involved tabulations of the number of "Don't Know" responses. From a theoretical standpoint it should be expected that with the least complex presentation of Stage I it should be easier to make a response than with the more complex, but not more meaningfully structured, presentation of Stage III. A weighting of 3, 2, and 1 was adopted for responses to Stages I, II, and III, respectively.

It might be mentioned that "Don't Know" responses for the meaningful stimuli also occurred occasionally. In this case, since that was already a wrong response and weighted accordingly, the rejections here were given one extra point to take into account this special type of response.

We had, thus, a quantitative measure of what might be called both perceptual distortions and rejections. In addition, the following five qualitative types of responses, each given a weight of 5, were considered as being both among the most significant and, as another criterion for selection, the most easily and objectively identified:

1. *Bizarre, contaminated, or symbolic responses.* These were responses in which the content was flagrantly unreal, involved the combination of two incompatible items or utilized the content as symbolic of great personal significance.

2. *Sex responses.* These were responses in which the drawing included the identification of some primary or secondary sexual characteristic.

3. *Morbid or horror responses.* These were responses in which the content included some dysphoric feeling tone.

4. *Cut-off responses.* These were responses which involved some violation of the objective limitations of the stimulus material and ignored certain parts of the figure presented, changed parts of the figure, or indicated that the drawing should continue outside of the card.

5. *Repitition.* This scoring involved repetition of responses from one stage to another and from one item to another. This category was considered only for the nine nonsense figures. The reason was that for these items there was no meaningful development from stage to stage and therefore repetition was more correctly a perseveration than it was in the case of the other nine stimuli.

Total scores were computed for each of the 125 cases and a product-moment correlation was obtained between these weighted scores and the sum of the four ratings for each subject. This correlation of .47, is given in Table 3, as are the correlations between these total scores and I.Q. and between I. Q. and ratings.

TABLE 3. INTERCORRELATIONS AMONG TEST SCORES, RATINGS, AND I. Q.'S

Variables	r	σr
Rating and Test Score	.55	.063
Rating and I. Q.	−.38	.077
Test Score and I. Q.	−.39	.076
$r_{12.3}$	.47	.071

Using the split-half correlation as a measure of reliability and correcting for length by the Spearman-Brown formula, the reliability coefficient for the test scores was found to be .76. However, in correcting for attenuation the correction was made in the criterion only [6, p. 530]. The reliability for the ratings was applied to the first-order partial correlation given in Table 3 and the corrected correlation between the two measures was found to be .54.

Further examination of the relationship between the two variables was made by grouping the test scores according to the five rating groups. (The total score of the four pooled ratings was divided by the number of ratings to give an average rating which was then rounded off within the original five-point grouping.) Table 4 shows that, while there was considerable overlap of test scores for adjacent rating groups, there was a distinct trend toward differentiation of all five groups.

TABLE 4. DIFFERENTIATION OF TEST SCORES BY RATING

Rating	Mean	S. D.	σM	Range	P<
1.	20.00	6.69	1.46	8–37	.10
2.	23.64	6.83	1.49	12–38	.10
3.	27.43	10.06	1.73	13–56	.10
4.	32.69	12.94	2.54	15–61	.02
5.	47.11	20.86	4.92	24–99	

DISCUSSION

Sherif and Cantril, in their discussion of gradation of structure in the stimulus field [9, pp. 45-48], state that few studies have investigated this variable. This observation is borne out in our own review of the literature. However, they believe that even present meager results point to the fact that increased structuredness produces decreased influence of internal factors in the percept.

We have already noted that for the entire sample the proportion of incorrect responses to the meaningful stimuli decreases considerably from Stage I through Stage III (Table 2). It is these data which seem to verify the above hypothesis. When the organization of the stimulus material is such that its structure seems to compel immediate and unequivocal recognition, even severe internal pressures such as accompany extreme personality maladjustments do not have a major influence on the perception.

However, while it was expected that degree of structuration would have an important influence on the perception, we expected to find the operation of another variable, degree of personality integration, to be influential within this framework. It would seem that the extent to which the perception takes on attributes of the structure of the stimulus increases more rapidly, with changes in structuration in the stimulus, in well-adjusted individuals than in severely disturbed individuals.

Bruner and Postman [2, pp. 83-94] speak about *selection*, *organization*, *accentuation*, and *fixation*, as four broad functions performed by the perceiver. Subjective factors may influence the perception in any one or all of these functions. Furthermore, for each of these functions there are undoubtedly a number of different processes, in accord with specific personality dynamics, whereby the subjective factors may influence the perception.

We emphasize this point in order to help explain the extensive overlapping which occurred among the test scores for the five rating groups. The analysis of the stimulus material, while sufficient to confirm our original hypothesis, was not capable of revealing the full relationship of our two variables. Our test material tapped only a limited number of ways in which the internal factors might operate in the perceptual process. This observation explains not only the overlapping in scores among the ratings but also the marked increase in variability of scores among the severely maladjusted subjects. As we go from normals to the severely maladjusted individuals it would seem that an increasing amount of the personal factor is introduced into the perceptual process. If this is accompanied not only by increased intensity in the operation of the personal factor but by an increasing number of different methods of introducing these personal factors then any technique which is sensitive to only a constant number of methods would become less and less specific to the total process as an increasing amount of the personal factor operates. Thus, for our Rating 1, where the personal factor, for all three stages, is least operative, there would be lowest scores no matter if only a portion (but of necessity a significant portion) of the methods was measured. On the other hand, as the intensity and diversity of subjective factors increased, some scores would be high—where the methods used were those tapped by the technique; and some scores would be deceptively low— where the methods used were mainly outside of those measured by the technique.

Further Results

In an attempt at some cross validation of the results found here, the graded stimulus material was presented to 68 additional out-patients at a Veterans Administration mental hygiene clinic. Again, these were white, male veterans with no known organic brain damage. The scores on the test were correlated with a single rating on the five-point rating scale. This rating was made by the psychologist doing the complete psychological examination. Two changes in the original procedure were made. Since this sample of 68 cases varied less in severity of illness than the original group, half step intervals were used in the rating to increase the number of groupings from three to six. The range of ratings went from 2.0 to 4.5 with a mean rating of 3.1. The stimulus material was unchanged in content and scoring except that three of the series; the tree, the watch, and the profile were omitted from each stage. These three had been among the least discriminating of the original 12 items. Thus, there were 27 rather than 36 cards in all. The product-moment correlation was .52, which is quite close to the uncorrected correlation of .55 on the original 125 cases and would seem to argue that our a priori analysis and scoring did achieve some measure of relationship to the rating of personality integration.

Summary and Conclusions

This study attempts to investigate one aspect of the very complex relationship between personality and the perceptual process. Degree of personality integration was measured by using a five-step rating scale on 125 subjects obtained from varying sources. Reliability of these ratings, established by intercorrelation of three independent ratings, was .75.

Structuredness of stimulus was graded into three steps by taking a series of nine incomplete line drawings of simple objects and removing, in two successive steps, key details of the drawing. In addition to the nine drawings, three nonsense drawings were interspersed so as to prevent a mental set from developing in regard to the

increasing structuredness from stage to stage. Responses to the graded stimulus test were analyzed and scored according to three aspects: failure to see the "expected" response, rejections, and various qualitative indications. Total weighted scores on this test were then correlated with the Mean Ratings for the subjects and the correlation of .55 proved substantially significant. However, intelligence, measured by the Wechsler-Bellevue Scale, was also found to correlate significantly with both ratings and test scores and, therefore, had to be partialled out. The partial coefficient between ratings and test scores was .47.

The results did largely confirm our original hypotheses and the following conclusions seem warranted in the light of this study:

1. All other things being equal, the role of internal factors in the perceptual process decreases with increased structuration of stimulus.

2. All other things being equal, the role of internal factors in the perceptual process increases with decreased total personality integration. As a corollary of this it might be stated that individuals who are well-adjusted are quicker to restructure their percepts in accord with the stimulus as structuredness increases than are individuals with severe maladjustments.

The implication of these findings is that the relationship between the perceptual process and the level of personality integration, although complex, is distinct and substantial.

References

1. Bruner, J. S., and Goodman, C. C. Value and need as organizing factors in perception. *J. abnorm. soc. Psychol.*, 1947, 42, 33-44.
2. Bruner, J. S., and Postman, L. An approach to social perception. In Dennis, Wayne (Ed.), *Current trends in social psychology*. Pittsburgh: University of Pittsburgh Press, 1948, pp. 71-118.
3. Bruner, J. S., and Postman, L. Perception, cognition and behavior. *J. Personality*, 1949, 18, 14-32.
4. Buhler, C., Buhler, K., and Lefever, D. W. *Rorschach Standardization Studies: 1. Development of the Basic Rorschach Score.* (Rev.) Los Angeles: Rorschach Standardization Studies, 1948.
5. Frenkel-Brunswik, Else. Intolerance of ambiguity as an emotional and perceptual variable. *J. Personality*, 1949, 18, 108-143.
6. Guilford, J. P. *Fundamental statistics in psychology and education.* (2nd. Ed.) New York: McGraw-Hill, 1950.
7. Klein, G. S., and Schlesinger, H. Where is the perceiver in perceptual theory? *J. Personality*, 1949, 18, 32-48.
8. McNemar, Quinn. *Psychological Statistics.* New York: Wiley, 1949.
9. Sherif, M., and Cantril, H. *The psychology of Ego-involvements.* New York: Wiley, 1947.

AN ANALYSIS OF THE RATIONALIZATION AND EXPLANATIONS OF TEST BEHAVIOR

JOSEPH E. SHORR, PH.D.

477 South Fairfax Avenue
Los Angeles 36, California

Problem

The process of rationalization as an ego defense mechanism has been universally recognized. Since Jones first used the term in 1908 there has been a great deal of descriptive literature about rationalization. However, there exists a paucity of experimental or statistical studies centered about this self-justification concept. This is especially true with respect to the commonly heard rationalizations or explanations offered by subjects for not doing as well as they might have done on paper and pencil tests usually given in vocational guidance.

Because of its frequent occurrence in everyday life, an analysis of rationalization has implications for the psychotherapist, teacher, parent, and all personnel workers who are interested in the ego-defenses of the people they deal with.

The purpose of this study was (1) to categorize the reasons offered to explain test behavior and determine their frequency; (2) to compare the categories of reasons with each other with reference to objective test scores of temperament and the abilities.

Method

Preliminary survey. In order to gather as many rationalizations as possible, a random portion of a senior high school class was canvassed to obtain the one most important reason why they did not do themselves justice on any or all tests they had taken in the past. No tests were administered previous to this request. They responded to a prepared statement reading verbatim as follows:

> "We often hear people say that they could have done better in tests than they have done. We would like for you to write the one most important reason why you think you could have done better than you have."

The rationalizations or reasons thus gathered were given to a committee of three psychologists who were instructed to label each reason with a word, a phrase or sentence that would in their opinion characterize the statement. The judges worked independently and labeled each statement in any fashion they chose.

Inspection of the judges' statements which overlapped and showed agreement revealed the following clusters of reasons:

1. Time or speed factors;
2. Nervousness or anxiety;
3. Carelessness or poor work habits;
4. Self-depreciation or inferiority feelings;
5. Poor motivation or lack of concern;
6. Did best or no excuse;
7. Body aches or physical discomfort;
8. Tests were inadequate to measure what they were supposed to;
9. Never got started or misunderstood directions;
10. Miscellaneous.

Since the great majority of the reasons for test behavior secured were direct and clear as to meaning, it was decided by the committee of judges that only those statements which had unanimous agreement would be acceptable for categorization. Of the 137 statements in the preliminary survey, six were eliminated because of being illegible, blank, or semantically unclear, leaving 131 statements on which there was perfect agreement.

From these a set of directions for selecting categories of rationalization was developed, and this became the basis for subsequent categorization. These are as follows:

Directions for Selecting Categories of Rationalization

Each card containing a rationalization will be numbered. When you have chosen a category, say Rationalization 1, Rationalization 2, etc., for each card, first recording the card number and then the Rationalization number on the sheet provided.

Read the following examples for each category before selecting

Rationalization 1: When a doubt is expressed as to time, as "test is too short", "worked too slowly", etc.

Rationalization 2: When tenseness or emotion is expressed, as "I was tense", "I was nervous", etc.; expressions of instability, moodiness, being bothered by noise, etc.

Rationalization 3: When careless work habits, poor study methods, or working on the wrong area of the answer sheet is expressed, etc.

Rationalization 4: When self-depreciation is expressed, as "Never do well on tests", or "I shouldn't be taking tests", etc.

Rationalization 5: When poor motivation is expressed, as "don't care", "it makes no difference", etc.

Rationalization 6: When no excuse, or "I think I've done my best", is expressed.

Rationalization 7: When bodily aches or pains are mentioned: headache, cold, complaints about eyes, glasses, light, hearing, weather, etc.
Rationalization 8: When the test is considered unfair or inadequate for measuring the individual's particular interests, abilities, etc.
Rationalization 9: When instructions are misunderstood and the individual did not know what to do.
Rationalization 10: When none of the above rationalizations is expressed.

Following the delineation of the rationalization categories to be used in this study, specific statements from students justifying their test behavior after certain standardized tests of temperament and mental abilities had been administered were secured. The standardized tests used were the PMA, the Guilford-Martin GAMIN-STDCR factors, the Memory Test of the Factored Aptitude Series developed by King, and the Perceptual Speed test of the Guilford-Zimmerman Aptitude Survey.

Subjects. Five hundred sixty-one high school seniors comprised the sample utilized in the principal portion of the investigation. Data were obtained on seven hundred twenty-four students, but one hundred sixty-three cases were eliminated because administrative rules prevented the use of the test scores of one of the schools. The rationalizations of these one hundred sixty-three students were kept, however, for frequency comparisons.

Procedure. After the standardized tests had been administered, the prepared request to justify their test behavior was read to them as used in the preliminary survey. The explanations or rationalizations were then categorized by the use of the Directions for Raters on Selecting Categories of Rationalization. The same three judges participated in the categorizing of those statements as were utilized in the preliminary survey.

Results

Of the 724 statements so categorized, there was not a single case of rationalization 4 or "self-depreciation", rationalization 8 or "test unfair", or rationalization 9 or "instructions misunderstood". To nullify ambiguity, rationalization 10 or "miscellaneous" was eliminated from the investigation. This left the following six rationalization categories: (1) time, (2) nervousness, (3) carelessness, (5) poor motivation, (6) did best or no excuse, and (7) body aches and pains.

The procedure described above was followed in three separate schools of similar population in order to check the consistency of the frequency of the various categories of test behavior explanation or rationalizations. It can be seen from Table 1 that the percentage of rationalizations secured in the three high schools shows ex-

Table 1. Percentages of Types of Rationalizations Secured from Three High School Student Populations

Rationalization Category	School A*		School B		School C	
	N	Per cent	N	Per cent	N	Per cent
1. Time	80	49	132	41	112	48
2. Nervousness	30	18	74	23	40	17
3. Carelessness	26	16	58	18	47	19
5. Poor Motivation	2	1	4	1	5	1
6. Did Best	8	5	18	6	6	3
7. Body Aches	17	11	38	11	27	12
Total 724	163	100	324	100	237	100

*This school was used only to check frequency of responses.

tremely high consistency. "Not enough time" was given as a rationalization to the extent of about forty-five per cent of the total reasons. "Nervousness" as a justification was presented to the extent of approximately nineteen per cent of the total rationalizations secured. Approximately eighteen per cent of the total reasons secured were categorized as "carelessness". Only one per cent of the student population gave a rationalization of "poor motivation". Those who claimed they had "done their best" made up about four per cent of the total. Finally, those who used "body aches and pains" as a reason to justify their test behavior made up eleven per cent of the total sample.

The two high school senior classes used in this study were compared as to the frequency of rationalizations expressed by the two sexes. As Table 2 indicates, fifty-seven per cent of the total populations were females and forty-three per cent were males. The "time" group showed approximately the same percentage distribution as the total distribution (critical ratio of percentages .80). The "nervousness" group showed somewhat less similarity, but yet were not significantly different (critical ratio of percentages .99). The distribution showed that the males used "carelessness" as a rationalization significantly more often than the females based upon the expected expectancy total distribution (critical ratio of percentages 1.89 or approaching significance at the five per cent level). The male "poor motivation" group

TABLE 2. PERCENTAGES OF TYPES OF RATIONALIZATIONS BY SEX FROM TWO HIGH SCHOOLS

Rationalization Category	Male		Female		Total	
	N	Per cent	N	Per cent	N	Per cent
1. Time	99	40	145	60	244	100
2. Nervousness	43	38	71	62	114	100
3. Carelessness	55	53	50	47	105	100
5. Poor Motivation	5	56	4	44	9	100
6. Did Best	12	50	12	50	24	100
7. Body Aches	28	43	37	57	65	100
Total	242	43	319	57	561	100

did not occur significantly more often then the "poor motivation" explanation for the females with reference to the total distribution (critical ratio of percentages .78). The male "did best" group did not occur significantly more often than the female with reference to the total distribution (critical ratio of percentages .67). The "body aches" group coincided in percentage distribution with the total distribution of boys and girls (critical ratio of percentages .00).

COMPARISON WITH OTHER TEST RESULTS

The significant differences between each of the rationalization groups and all the other rationalization groups combined on the temperament and mental abilities factors were determined in order to yield a particular pattern (or lack of pattern) for each category.

The "time" group. This group of students, who complained about the pressure of time while taking tests, showed themselves to be significantly less active in their general overt behavior than all the other rationalization groups combined. The "time" group rated itself as being significantly less stable and more introverted in its thinking than all the groups combined. In other words, the last stable students used "time" as a reason for not doing well on examinations. Moreover, the "time" group was significantly inferior on its earned scores on the verbal, spatial, reasoning and word-fluency factors than all the other groups combined. In general, the less stable,

the less active, the more meditative and the poorer students in the mental abilities offer pressure of time as an excuse for not doing as well as they could have done when taking tests.

The "nervousness" group. The group who gave "nervousness" as a reason for poor test behavior show no homogeneity when compared to all the other groups combined. No distinct patterns are discernable from the data. Students who gave "nervousness" as a rationalization for not doing well when taking tests indicated no significant characteristics on any of the seventeen factors utilized in this study.

The "carelessness" group. The group who gave "carelessness" as a reason for inadequacy in taking tests indicate they are more "masculine" in attitude and interest than all other rationalization groups combined. The data suggest that those students who are more "feminine" than others in attitudes, tend not to rationalize test behavior by offering "carelessness" as an excuse. Since, as Table 2 indicates, significantly more males than female students manifested this rationalization, superiority in space perception for this category may be expected in view of the fact that in our culture boys are generally conceded to be "better" in spatial conceptions than girls. This "carelessness" group is superior to all the others in number ability and in word-fluency. In general, the group that uses the rationalization "carelessness" rates itself more masculine and scores higher on the space, number and word-fluency factors than all the other groups combined.

The "poor motivation" group. This group, which was small in total percentage (see Table 2), nevertheless showed certain definite differences from all the other rationalization groups combined. Despite the group's "don't care" or "it makes no difference" approach, the group rated itself as being introvertive in thought and decidedly lacking in fluency with words in comparison with all the other groups combined. Furthermore, this group was less able than the others to discern details speedily. The rationalization stated may reflect poorer communication ability on the part of this group than all the others combined. The data suggesting less "care-

TABLE 3. CRITICAL RATIOS BETWEEN EACH RATIONALIZATION GROUP AND ALL OF THE OTHER RATIONALIZATION GROUPS COMBINED ON THE VARIOUS TEST FACTORS

Test Factor	Critical Ratios					
	Time	Nervousness	Carelessness	Poor Motivation	Did Best	Body Aches
Sociability	1.27	0.57	0.09	0.17	1.15	1.77 # # #
Meditativeness	2.02 # #	0.83	0.24	2.89 #	1.38	3.72*
Depression	1.79 # # #	0.25	0.66	0.68	0.63	1.85***
Cycloid	2.13 # #	0.59	0.59	0.74	0.28	2.23*
Rhathymia	0.73	0.79	0.42	0.02	0.15	0.34
Activity	3.06 #	0.57	0.42	0.23	0.88	3.35*
Assertiveness	1.05	0.55	0.15	0.71	0.35	2.08**
Masc.-Femin.	0.38	0.84	2.48*	0.17	0.54	0.58
Inferiority	0.02	1.41	1.52	0.16	0.42	1.19
Nervousness	1.45	0.80	0.72	0.34	0.68	1.31
Verbal	2.11 # #	0.74	1.19	1.47	0.32	1.66***
Space	2.47 #	0.80	1.75***	1.16	1.38	0.10
Reasoning	1.82 # # #	0.47	0.66	1.18	0.63	0.10
Number	0.34	0.87	2.85*	1.22	1.23	0.69
Word-Fluency	1.69 # # #	0.44	1.92***	1.96 # #	0.44	1.43
Memory	1.10	0.93	0.19	1.59	0.61	1.25
Perceptual Speed	1.22	1.50	0.64	2.34 #	0.23	1.05

\# / # # / # # # } Group so designated scores significantly lower than the other groups combined, at the { one / five / ten } percent level of confidence, respectively.

* / ** / *** } Group so designated scores significantly higher than the other groups combined, at the { one / five / ten } percent level of confidence, respectively.

lessness" in thought may explain the lack of speed in discerning perceptual details. In general, this group is introvertive in thought and inadequate in word-fluency and perceptual speed.

The "did best" group. This group is apparently heterogeneous in composition. Students using this rationalization showed no distinction of statistical significance from all the other rationalization groups combined.

The "body aches" group. This group shows more temperamental differences than any of the other groups. The "complainer" of various body aches and pains indicates poorer sociability than all of the other rationalization groups combined. The members of this group are more emotionally stable and show greater aggressiveness than those in the other categories. Moreover, they are decidedly less introvertive in thought and more active overtly than the others. As far as the mental abilities are concerned, the members of this group are superior in word-meaning or verbal ability. The data suggest that, because of their superior activity rate, aggressiveness, "carefreeness" of thought, and verbal facility, members of this group may comprise many of the individuals who are "troublesome" to test administrators.

Conclusions and Summary

The results of the present study suggest that explanations or rationalizations of test behavior are amenable to categorization and statistical manipulation. Furthermore, a high degree of consistency exists in terms of the frequency of percentages of the kinds of rationalizations offered by students of the schools studied in this research.

The only sex difference noted in frequency of rationalizations used was "carelessness", which was used significantly more often by boys than girls. This may have a cultural basis.

The "time" group was less stable, less active, more introvertive in thought and poorer in the verbal, spatial, reasoning and word-fluency factors of mental ability than all the other categories combined.

The "nervousness" group showed no distinct pattern when compared to all the other groups combined.

The "carelessness" group are more masculine and superior in spatial, numerical and word-fluency factors of mental ability.

The "poor motivation" group showed poorer word-fluency and perceptual speed and is more introvertive in thought despite its "don't care" approach.

The "did best or no excuse" group indicated no significant differences from all the other groups combined.

The "body aches" group showed poorer sociability, greater emotional stability, greater aggressiveness, less introversion in thought, a higher activity rate plus better verbal ability than all the other groups combined.

Bibliography

1. Coleman, James. *Abnormal psychology and modern life.* Chicago: Scott, Foresman and Company, 1950.
2. Garrett, Henry E. *Statistics in psychology and education.* New York: Longmans, Green and Company, 1947, pp. 190-191.
3. Guilford, J. P. *An inventory of factors S T D C R.* Beverly Hills, California: Sheridan Supply Company, 1948.
4. Guilford, J. P. and Martin, H. G. *The Guilford-Martin inventory of factors G A M I N.* Beverly Hills, California: Sheridan Supply Company, 1948.
5. Guilford, J. P. and Zimmerman, W. *Aptitude survey, manual of instructions and interpretations.* Beverly Hills, California: Sheridan Supply Company, 1947.
6. Jones, Ernest. Rationalization in everyday life. *J. abn. soc. Psychol.*, 1908, 3, 161-169.
7. King, J. *Handbook for use of the factored aptitude series.* Chicago: Industrial psychology, 1948.
8. Sears, Robert R. *Experimental analysis of psychoanalytic phenomena.* Pp. 306-332 in J. McV. Hunt: *Personality and the behavior disorders,* Vol. 1. New York: The Ronald Press.
9. Symonds, Percival. *The dynamics of human adjustment.* New York: Appleton-Century Crofts, Inc., 1946.
10. Thorpe Louis. *Child psychology and development.* New York: Ronald Press, 1950.
11. Thurstone, L. L. and Thurstone, T. G. *Examiner manual for the SRA primary mental abilities.* Chicago: Science Research Associates, 1946.

A QUANTITATIVE STUDY OF THE HTP AND ITS RELATIONSHIP TO THE WECHSLER-BELLEVUE SCALE[1, 2]

HAROLD RUBIN

Veterans Administration Hospital, Lebanon, Pa.

INTRODUCTION

During the past two years, this writer has used the House-Tree-Person Test (HTP) devised by Buck [1] as part of a routine admission test battery in a veterans hospital. Although devised as a projective device aimed at appraising total personality structure, the HTP does have a carefully designed quantitative procedure which yields a number of IQ scores. While in one section of his manual Buck states that the HTP is not intended to be a test of "intelligence" and that no high correlations should be expected between it and other specific measures such as the Wechsler-Bellevue or Stanford-Binet, he claims in a later section that the test does measure "general intelligence". An apparent inconsistency exists. Furthermore, Buck presents tables showing rather high correlations between the HTP and W-B scores. This writer is of the opinion that Buck may be underestimating the possible effectiveness of the HTP as a measure of intellectual function in the interest of furthering its value as a projective device measuring total personality structure.

Estimating intellectual level from drawings is certainly no new approach, but it became a more widely utilized method with the advent of the Goodenough Draw-a-Man-Test in 1926 [4]. The method was, for many years, confined largely to drawings of children and it was only recently that this approach became more widely used at the higher age levels. The HTP is one of these more recent approaches and it is the aim of the present investigation to explore its value as a measure of adult intelligence.

The only report this author could find in the literature evaluating the HTP as a measure of intellectual function is that of Sloan and Guertin [6] in a study on mental defectives. These authors fail to find evidence for support of the hypothesis that the HTP is a measure of intelligence using the W-B as the criterion. These findings, however, represent a very limited population. The present study attempts to evaluate this hypothesis with a wider range of subjects and with the additional aim of establishing its validity in diagnostic screening. For this purpose, the Wechsler-Bellevue Scale was selected as the most widely accepted measure of that function which we call "intelligence."

PROCEDURE

Subjects in the present study initially consisted of 135 patients admitted to the N. P. Service of a veterans hospital. Although specific diagnosis was not a criterion for selection, many of the subjects were psychotic. All testing was done within approximately one month following admission to the hospital and included the HTP and Wechsler-Bellevue tests among others. Of this original group of subjects, twenty-seven were eliminated from the final group because their HTP's were regarded as non-scorable for this study. They either failed to meet the minimal requirements for scoring as outlined by Buck [1] or they produced incomplete records by omitting one of the three drawings. The Wechsler-Bellevue tests for these subjects, however, were scorable. Although this procedure may be regarded as somewhat selective, it was felt that the nature of the investigation supported this approach. The remaining 108 records were then utilized in this study. The subjects ranged in age from 18-50 years with a mean of 30.2 and in educational level from four years to college graduate with a mean of ten years. Standard administrative and scoring procedures were followed at all times as outlined by Buck [1].

---

[1] Read in part at the meetings of the American Psychological Association, Cleveland, O. 1953.

[2] From the Veterans Administration Hospital, Lebanon, Pa.

Results and Discussion

In Table 1 are presented the mean scores, standard deviations and ranges for the HTP and W. B. IQ's. It can be readily noted that both the *% Raw G* IQ and *Net Weighted* IQ scores of the HTP tend to be somewhat lower than either of the three WB IQ scores and also that the former have a slightly wider dispersion. An F test of the five mean scores failed to yield a significant difference.

Table 1. Means, Standard Deviations and Ranges of HTP and Wechsler-Bellevue IQ's (N = 108)

Test	Mean	S. D.	Range
HTP % Raw G IQ	95.18	21.2	36 – 140
HTP Net Wtd. IQ	92.15	19.7	35 – 134
WBFS IQ	101.58	16.5	55 – 127
WBV IQ	100.29	17.6	56 – 129
WBP IQ	102.56	14.9	62 – 133

Table 2. Pearsonian *r*'s Between HTP % Raw G, HTP Net Weighted IQ and Wechsler-Bellevue IQ's (N = 108)

Wechsler-Bellevue IQ's	HTP % Raw G	Confidence limits (P=.99)	HTP Net Weighted IQ	Confidence limits (P=.99)
Verbal	.623	.44 – .75	.604	.42 – .74
Performance	.634	.46 – .76	.580	.39 – .72
Full Scale	.674	.51 – .79	.636	.46 – .76

Table 2 presents the correlations between the W. B. IQ scores and the HTP *% Raw G* IQ scores. The confidence limits listed were established by means of the *r* to *z* transformation described in McNemar [5]. As can be observed, the *% Raw G* IQ correlates highest with the W. B. Full Scale IQ (r = .674), next highest with the W. B. Performance Scale (r = .634) and lowest with the W. B. Verbal Scale (r = .623). The difference between these three sets of correlations are actually very small. These correlations, although not quite as high, compare favorably with those reported by Buck [1] on his standardization group. Buck found correlations between % Raw G IQ's and WB IQ's to be .746 for the Full Scale, .724 for the Performance Scale and .699 for the Verbal Scale.

The *Net Weighted* IQ was treated in a similar manner and the results are given in Table 2. Correlations between the WB IQ's and the *Net Weighted* IQ were found to be .636 for the WB Full Scale, .604 for the WB Verbal Scale and .580 for the WB Performance Scale. These correlations are somewhat lower than those reported for *% Raw G* IQ and hence of lesser value in predicting W. B. scores.

These results would seem to indicate that the HTP does compare favorably with the Wechsler as a measure of intelligence thus contradicting the conclusions of Sloan and Guertin [6]. Some validity, then, may be attributed to Buck's contention that the HTP is a measure of intelligence, at least to the extent that it is measured by the Wechsler-Bellevue Scale. These findings indicate that in a screening situation one may obtain a fair estimate of a subject's level of intellectual function from the HTP. A more thorough evaluation, of course, would be obtained through administration of the Wechsler-Bellevue.

To evaluate the extent to which the age correction factor on the Wechsler affects the relationship between W. B. and HTP IQ scores, the latter having no such

"correction", a correlation was computed between the Raw G IQ and Full Scale Weighted Scores. An r of .534 was obtained which is lower than that obtained between Raw G and W. B. IQ scores (r = .674). This would appear to contra-indicate Buck's presumption that the relationship between the HTP and Wechsler-Bellevue IQ's would be depressed because of the age correction factor on the Wechsler. It does suggest, however, that the HTP score is not affected by advanced age to any appreciable degree and as such, adds to the stability of this measure. A comment might be made at this point that the HTP Raw G IQ bears a higher relationship to the WB Full Scale Weighted Score than at least three of the WB subtests: Digit Span, Picture Arrangement and Object Assembly.

To further evaluate the quantitative scores of the HTP, all cases in which there was a disparity of 15 points or more between HTP and Wechsler IQ's were explored to determine whether this difference was in some way related to intellectual level as established by the W-B. Fifty-three cases were found to reveal such a discrepancy. Of these, as observed in Table 3, sixteen were in the upper intellectual range, twenty-two in the middle range and fifteen in the lower range. This distribution tends to approximate the distribution of all subjects at these three intellectual ranges. It may be observed that in large part the differences were in the direction of lower HTP scores. Where HTP scores were higher than the Wechsler, they appeared in greater frequency at the lower intellectual levels. Sloan and Guertin [6] report a somewhat similar finding with mental defectives.

TABLE 3. FREQUENCY OF CASES WITH 15 OR MORE POINTS DISPARITY BETWEEN HTP AND WECHSLER-BELLEVUE IQ'S AT THREE INTELLECTUAL LEVELS

Wechsler-Bellevue	HTP Higher	HTP Lower	Total
115 and above	3	13	16 (30%)
91 – 114	3	19	22 (42%)
90 and below	5	10	15 (28%)

A further analysis of the data in Table 3 revealed that in almost all instances the % *Raw G* IQ more closely approximated the W. B. Full Scale score than did the *Net Weighted* IQ, irrespective of intellectual level. At the lower levels in particular, the weighted scores appeared to penalize the subjects heavily. This finding is in disagreement with Buck's contention that in subjects of above average or superior ability the Net Weighted IQ more closely approximates the subject's level of intellectual function. It should be added, however, that there is one exception to this. In those records in which no "flaw" scores appear, the percentage method of arriving at the % *Raw G* IQ yields a spuriously high score and in such instances, the Net Weighted IQ is a more accurate estimate of intellectual function. However, as Buck has indicated, the frequency of this occurrence is rare, with only two such records appearing in this study.

More recently, Buck [3] has attempted to develop score patterns indicating maladjustment based on what he terms "the quality of the quantity." One such pattern involves a comparison between Raw G and Net Weighted IQ's. According to Buck, "The Raw G IQ . . . seems largely to represent the quantity of a subject's creativity and productivity; to emphasize how *much* he does. The Net Weighted IQ . . . seems to emphasize the quality of the subject's concepts; to stress how *well* the subject does something." A discrepancy of fifteen or more points in favor of the Net Weighted IQ is regarded as indicating that the subject is functioning below the qualitative level of which he is capable. On the other hand, if the % *Raw G* IQ is 15 or more points above the Net Weighted IQ, "pathoformicity is definitely to be suspected, and with the quality of the behavior that much below quantity, regression is implied. . ." Comparisons between the Raw G and Net Weighted IQ's of the 108 subjects studied

revealed that only 7 had differences of fifteen or more points, all in the direction of the Raw G score. Differences of ten or more points were found in nineteen patients with only 2 in the direction of the Net Weighted IQ. The mean difference between Raw G and Net Weighted IQ's was 6.8. Since the subjects in this study were all psychiatric patients, it would be expected that if this pattern were valid more than seven patients should have been identified by this discrepancy. Thus, these findings question the validity of such a pattern analysis with the HTP.

Summary

The HTP was evaluated with respect to its effectiveness as a measure of intellectual function. Subjects were 108 patients admitted to the N. P. Service of a veterans hospital. With the Wechsler-Bellevue as the criterion, results revealed the HTP to be a valid measure of "intelligence." The % *Raw G* IQ had the greatest value in this respect. Buck's score pattern proposed to indicate maladjustment did not identify this group of psychiatric patients as "maladjusted."

References

1. Buck, J. N. The H-T-P technique. A qualitative and quantitative scoring manual. *J. clin. Psychol. Monogr. Suppl.*, 1948, No. 5.
2. Buck, J. N. Administration and interpretation of the HTP test. *Mimeo. manual*, Western Psychol. Services, 1950.
3. Buck, J. N. The quality of the quantity. *J. clin. Psychol.*, 1951, 7, 352-356.
4. Goodenough, Florence L. & Harris, Dale B. Studies in the psychology of children's drawings. II. 1928-1949. *Psychol. Bull.*, 1950, 47, 369-433.
5. McNemar, Q. *Psychological statistics.* New York: Wiley, 1949.
6. Sloan, W. & Guertin, W. H. A comparison of HTP and Wechsler-Bellevue IQ's in mental defectives. *J. clin. Psychol.*, 1948, 4, 424-426.
7. Wechsler, D. *The measurement of adult intelligence.* (3rd Ed.) Baltimore: Williams & Wilkins, 1944.

THE USE OF CARTOONS AS A PROJECTIVE DEVICE

GEORGE B. STROTHER, MARGARET MCCRACKIN BARNETT AND PETER C. APOSTOLAKOS

University of Minnesota, Duluth Branch

Introduction

Many psychologists accept Freud's theory that comic situations owe much of their effect to the fact that they release repressed reactions in a socially acceptable way. The authors were impressed with the extent to which this seemed to be true of the series of cartoons by William Steig published under the title *The Lonely Ones.* (2) The subsequent appearance of these cartoons on cocktail napkins, ashtrays and cocktail glasses may, through wider dissemination, have reduced their potential value as a clinical instrument but it has at the same time assured their use under conditions whereby maximum catharsis might be obtained. Since the experimental trials for the present study antedated the use of these cartoons as an adjunct to the cocktail hour we need not be concerned with the effect thereof as far as this study goes.

Purpose

The purpose of this study was to determine whether such cartoons as Steig's might be used as the basis for a projective technique. As a means of determining the subject's reactions to the cartoons, several possibilities were considered. Finally a preferential sorting was chosen as the simplest means of determining individual reactions.

Methods

Preliminary studies were done with 60 subjects, all college students, comparing two different sorting procedures. Both procedures involved having the subjects sort the cartoons into five categories. The first procedure involved sorting them into "like-dislike" categories and the second into "funny-not funny" categories. The cartoons had been removed from the book and mounted on stiff cardboard with the captions below. Each was numbered, the numbers corresponding to their serial order in the book. These first runs were all individually administered. The question was also raised whether the order of presentation might have some effect and runs were tried in a 1 to 46 order and in a 46 to 1 order. No difference was found in the tendency of subjects to use any one category systematically as a function of the order of presentation.

The cartoons were also presented to a group of 101 subjects by means of an opaque projector and responses were recorded by the subjects on an answer sheet. Subjects were instructed not to laugh at or comment on the cartoons so that other members of the group would not be influenced by their reactions. As a result of these preliminary runs, it was decided to use group presentation and to use the "funny-not funny" dimension. This was selected because it appeared to give subjects a feeling of making objective judgments whereas the preferential sorting by likes and dislikes was construed as a more personal matter. The five categories employed were as follows: (1) very funny, (2) funny, (3) indifferent, (4) not funny, and (5) disgusting. All subjects seemed to accept these categories and to use them as points on a scale. Responses were recorded by each subject on an answer sheet with the five choices in a row following the item numbers. Subjects were instructed to look at each cartoon and classify it into one of the five possible categories. Thirty seconds proved to be sufficient exposure for each cartoon.

The subjects used in this part of the experiment were 101 male and female college students enrolled in an introductory psychology course. This group of students had taken several tests and several of these were selected as possible criteria. These were (1) nine scales of the Minnesota Multiphasic Personality Inventory, (2) ratings of tendency toward psychosomatic illness based on evaluation of student health records, (3) the Schneider "index of autonomic efficiency"[3], and (4) scores of the "Reporting Test"[1, 3]. The psychosomatic ratings were based on a five point scale: a score of one indicated a health record with no indication of psychosomatic illness, a score of five indicated the presence of a clearcut chronic complaint such as recurrent bronchial asthma or several less severe but recurrent disturbances such as hives, hay fever and digestive complaints. The "Reporting Test" is a group frustration test described in detail in the articles previously referenced. High scores on this test indicate minimal disruption of test performance following frustration.

MMPI scores were available on all the 101 subjects. Psychosomatic ratings, reporting test scores and Schneider indices were available on 90, 52 and 58 subjects respectively.

The procedure for determining the relationship of cartoon reactions to each of the criteria was to compare the top 30% of the subjects with the bottom 30% for each of the criteria except the reporting test and Schneider groups wherein comparisons were made between the high and low halves of the groups. These high and low criterion groups were compared with respect to their distribution in each of the five choice categories; chi-square tests were applied to test the extent to which each of these pairs of distributions for each cartoon failed to meet the hypothesis of homogeneity.

Results

Table 1 lists each of the most significant cartoons for each of the criteria together with their probabilities of being true differences. Although it is not feasible to present the actual cartoons, reference to the captions will reveal that the cartoons showing maximum differences seem generally to be consonant with the supposed significance of the various criteria.

TABLE 1. LIST OF THE MOST SIGNIFICANT CARTOONS FOR EACH OF THE CRITERIA

Criterion	Cartoon No.	Caption	P
Hs	22	I can laugh at myself	.02
	38	Why pick on me	.05
D	14	Who are all those others	.02
Hy	1	I do what is expected of me	.05
	15	Mother loved me but she died	.10
	35	Let the leader of the cult do the worrying	.10
Mf	24	Public opinion no longer worries me	.10
	37	You've ruined my life	.10
Pa	17	Revenge is sweet	.10
	25	I do not forget to be angry	.10
	35	(see above)	.10
	46	But I can mystify and terrify	.10
Pt	37	(see above)	.10
Sc	37	(see above)	.10
Ma	9	My true love will come some day	.02
	14	(see above)	.05
	18	I have my own troubles	.10
Schneider index	10	I think before I act	.01
	13	Convention be damned	.10
	29	You'll have to take me as I am	.05
	30	I can't express it	.05
Reporting test	11	I am blameless	.05
	38	(see above)	.02
Psychosomatic rating	9	(see above)	.05
	33	I recreated myself	.01
Pd	15	(see above)	.10
	17	(see above)	.05

The fact that the probabilities in Table 1 do not often reach a high level of significance is not surprising. Single items of this type rarely would with samples of this size and with criteria whose validities are themselves problemmatical. Inspection of those items which do approach or reach a satisfactory level of significance suggests considerable face validity, particularly in view of the lack of any connection between the variables which might give a spurious relationship (such as intercorrelations between two inventories containing a number of similar or identical items). For example, the cartoons which differentiated subjects with high paranoid scores on the Multiphasic with the possible exception of number 35 are ones which most clinicians would be apt to rate as loaded with a paranoid connotation. The captions are self-explanatory in the cases of numbers 17 and 25. Number 46 shows a woman with very large eyes half concealed behind a screen on which a large eye is drawn.

Number 18, for another example not explained by the caption, rates high on the manic scale; it is a picture of an endomorphic man beating a large base drum.

In spite of the encouragement that the foregoing results gave the experimenters, it could not be said that the results thus far were by any means conclusive. The next step in the verification of the original hypothesis that certain types of cartoons might lend themselves to use as a projective device was to construct an experimental "test" based on the foregoing data. This was done in the following way. For each of the criteria a per cent difference between the high and low groups was computed for each of the five sorting categories. If the excess per cent favored the high group the sign of the difference was plus; if it favored the low group, minus. All differences were divided by 10 and rounded off to the nearest whole number. Thus a per cent difference of 4 received a value of zero, a difference of plus 27 a value of plus 3, etc. The values thus determined were considered as scoring weights.

The cartoons were then presented in group form to a new group of 31 subjects, also students in an introductory psychology course and their "scores" on each of the variables were computed. Table 2 gives the odd-even reliabilities corrected by the Spearman-Brown formula for each of the "scales". These reliabilities indicate that the subjects' reactions to the cartoons do tend to form stable clusters which in some cases approximate satisfactory test reliabilities.

TABLE 2. ODD-EVEN RELIABILITIES FOR THE EXPERIMENTAL CARTOON SCALES[1]

Hs	.67	Sc	.79
Hy	.82	Ma	.73
Pd	.80	D	.56
Mf	.29	Schneider	.43
Pa	.52	Reporting test	.41
Pt	.40	Psychosomatic rating	.59

[1]Corrected by the Spearman-Brown formula

An attempt was also made to determine whether these measures were reasonably independent of one another. Table 3 gives the intercorrelations between the various cartoon "scales". In some cases the intercorrelations are sufficiently high in view of the reliabilities to indicate that certain of the scales actually measure much the same thing. On the other hand some of the intercorrelations are low enough to indicate that several different factors are operating. Table 4, which is a table of intercorrelations between some scales of the Multiphasic for the original group of 101 subjects, indicates that the intercorrelations in Table 3 reflect to a large extent a lack of independence between the scales of the Multiphasic itself and do not necessarily inhere in the cartoons.

TABLE 3. INTERCORRELATIONS BETWEEN THE CARTOON "SCALES". (N 29)

Factors	Hs	Hy	Pd	Pa	Sc	Ma	Mf	Pt	D	Sch	R.T.	Pm
Hs		.73	.70	.63	.74	−.08	−.34	.55	.49	−.01	.17	−.01
Hy			.79	.84	.67	.07	−.57	.59	.48	.26	.06	.39
Pd				.76	.83	.04	−.13	.55	.49	.14	.19	.37
Pa					.69	−.20	−.33	.59	.15	.10	.20	.57
Sc						.01	−.32	.65	.19	.20	.04	.39
Ma							.17	.15	.04	.25	.09	−.01
Mf								.17	−.22	.53	.37	−.23
Pt									.26	.43	.41	.04
D										.37	.06	−.43
Sch											.51	−.05
R. T.												.42
Pm												

TABLE 4. TABLE OF INTERCORRELATIONS OF SELECTED SCALES OF THE MINNESOTA MULTIPHASIC (N 101)

Factors	Hy	Pd	Pa	Sc	Ma
Hs	.73	.70	.63	.74	−.08
Hy		.79	.84	.67	.07
Pd			.76	.83	.04
Pa				.69	−.20
Sc					.01
Ma					

As a further check on the consistency of the cartoons in predicting the criteria a small sample of 15 subjects was obtained on whom Minnesota Multiphasic scores were available. Rank difference correlations were computed on the cartoon scales vs. the corresponding sections of the Minnesota Multiphasic. Table 5 gives the values

TABLE 5. INTERCORRELATIONS BETWEEN CARTOON SCALES AND CORRESPONDING MMPI SCALES (N 15)

Hs	D	Hy	Pd	Mf	Pa	Pt	Sc	Ma
.52	.50	.43	.78	.61	.66	.46	.70	.81

obtained. These coefficients seem to indicate that the cartoon scales predict the criteria with a fairly high degree of accuracy. It seems reasonable to expect that were the other three criteria available on this small group similar correlations could be anticipated judging from the reliabilities obtained.

Blind analyses of cartoon scores have subsequently been undertaken with a small number of psychiatric patients and writeups based on these blind analyses have been matched with patients by a psychologist and psychiatrist who were familiar with the patients. The number of cases has been too small to reach a significant level of probability but appear to exceed chance expectancy.

DISCUSSION AND SUMMARY

The purpose of this study was to investigate the hypothesis that cartoons would prove useful as the basis for a projective technique. The cartoons selected for investigation were William Steig's *The Lonely Ones*, which appeared especially promising. It was found that certain cartoons differentiated between high— and low—scoring groups of subjects determined by the Minnesota Multiphasic, a rating of tendency toward psychosomatic illness, an "index of autonomic efficiency," and a test of response to a frustrating situation.

An experimental "test" was constructed based on these data. Per cent differences between the high and low groups of subjects on each criterion were used as the basis for determining scoring weights. The cartoons were then presented to a new group of subjects and reliabilities of the cartoon "scales" were computed. The results indicate that the subjects' responses to the cartoons tend to form internally consistent clusters which are in some cases more or less independent of one another. The interdependence found in other instances appears to reflect in part at least the interdependence of the criteria and does not necessarily reflect anything inherent in the cartoons.

The criteria selected have no particular virtue. They were chosen chiefly because of their availability and are justified solely on the basis of the original assumption that if such criteria showed positive results the original hypothesis would be verified. Other clinically or experimentally better criteria would almost certainly yield better results.

Item analysis and other refinements would doubtless improve the results obtained in the foregoing.

It is concluded that objectively scored judgments of cartoons have potential value as a projective technique.

BIBLIOGRAPHY

1. MCKINNEY, FRED; STROTHER, GEORGE B.; HINES, RUTH R. and ALLEE, RUTH A. Experimental frustration in a group test situation. *J. abn. soc. Psychol.*, 1951, 46, 316-323.
2. STEIG, WILLIAM. *The lonely ones.* New York: Duell, Sloan and Pearce, 1942, p. 104.
3. STROTHER, GEORGE B. and COOK, DONALD M. Neurocirculatory reactions and a group stress situation. *J. consult. Psychol.*, 1953, 17, 267-268.

A PRELIMINARY REPORT ON A TECHNIQUE DESIGNED TO DIFFERENTIATE PATIENTS WITH CEREBRAL PATHOLOGY AND PSYCHONEUROSIS*

ALEXANDER TOLOR

Neurological Institute, Presbyterian Hospital, N. Y. C.

Problem

Psychometric techniques that are able satisfactorily to differentiate between patients with cerebral abnormality and psychoneurosis are of great value particularly because of the similarity of symptomatology in these two categories. Also, the possibility of early detection in patients of a cerebral involvement before obvious clinical manifestations develop, and the possibility of altering the course of the disease by instituting early therapeutic measures, suggests that this area is an important one for further investigation. Of the numerous methods used by psychologists for neuro-psychiatric diagnosis, such as the analysis of performance on subtests of the Wechsler-Bellevue scale, the Rorschach inkblot test, projective drawings, and Bender-Gestalt Test, none is without serious limitations and deficiencies. The present report describes a very brief and economical instrument that may be easily adapted as one test in a psychological battery, in conjunction with other laboratory techniques, to assist in a differential diagnosis. The purpose of this investigation is to determine in what ways spontaneous colored chalk productions differentiate between patients with psychoneurosis and those with cerebral pathology.

Method

The materials used were white paper, size 8½" x 11", an ordinary stop watch, five colored chalks, blue, green, red, yellow, and black.[1] The paper was placed on the table directly in front of the patient with its longer axis oriented from left to right. The colored chalks were then arranged on the table in the following order: blue, green, red, yellow and black, reading from left to right. The examiner carefully named each color as he placed the corresponding chalk on the table in order to make certain that the patient became thoroughly familiar with all of the elements with which he was being confronted.

The following instructions were administered orally: "Using one or more of these colored chalks fill in the space on this paper completely. Make sure that you leave no white space. Remember you may do anything you like and use as many or as few of these colors as you wish as long as you leave no white space."

If the patient raised any questions they were answered in a manner consistent with the above instructions.

On his record sheet the examiner noted the name of each color as it was used and the time consumed for each from first contact with the paper to the return of the chalk to the table. Occasionally patients selected and disposed of chalks so rapidly that it was virtually impossible to obtain an accurate time measure, and in those cases a recording of "less than 5" was made. Two such "less than 5" recordings were taken to be equivalent to 5 seconds, and all time recordings were rounded off to the nearest 5 second interval.

In addition to these notations the examiner recorded various behavioral observations such as blushing, nail biting, humming, wiggling about in the seat, as well as any verbal comments made. The latter often reflected the patient's reactions toward specific colors yielding important clues concerning his expression of affect.

---

*The author wishes to express his indebtedness to Miss Louise Hewson, Director, Psychology Department, Neurological Institute, without whose guidance this work could not have been accomplished.

[1]All of the surfaces of the chalks were exposed and therefore available for use. Also all of the colored chalks were approximately equal in size, 3" having been found suitable, some allowance being made for differences in sizes resulting from the greater use of some of the chalks compared to others.

Discrepancies between the patient's announced plans and actual performance were recorded. Occasionally a patient requested other colors or permission to leave white areas undone. The examiner noted if there was an unusual degree of conflict manifested in selecting colors and observed how impulsive, deliberate, or hesitant the approach was. Other observations concerned the obliteration of certain colors, the pressure with which the colors were applied, the direction of approach, rotation of paper, and the degree of literalness with which instructions were followed. Some patients, for example, left considerable space undone, whereas others actually marked up the table top in an attempt to comply with the instructions. Further pertinent observations, similar to those the clinician makes when more conventional projective tests are administered, yielded significant clinical information. These incidental observations, although important for personality evaluation, were not subjected to analysis for the purposes of this investigation.

Subjects

The colored drawings examined for this study came from the productions of a group of adult patients seen at Neurological Institute, Columbia-Presbyterian Medical Center who were referred to the Psychology Department usually for the purpose of a differential diagnosis. The sample consisted of 38 patients, 21 later diagnosed as suffering from functional disorders (psychoneurotic group) and 17 later diagnosed as having a cerebral involvement. The criterion used was the final discharge diagnosis, whenever available, based upon a variety of laboratory techniques such as electroencephalogram findings, neurological examination, pneumoencephalogram results, skull X-rays, and psychological examination. Whenever discharge diagnoses were not available, as was the case especially with clinic patients, the criterion was the impression obtained from the psychological examination which included usually the Wechsler-Bellevue scale, Substitution Test, Bender-Gestalt Test, projective drawings, and Rorschach inkblot method. In the interpretation of the psychological findings the examiner relied heavily on the pattern of deviation ratios on the Wechsler-Bellevue Scale and the Substitution Test as they are used by Hewson [4] as aids in neuropsychiatric diagnosis.

The psychoneurotic group ranged in age from 16 to 60 years, mean 37.5 years (S. D. 10.6). Their range of intellectual functioning extended from the low average classification, an IQ of 86, to the superior classification, an IQ of 124. The mean quotient was 107 (S. D. 11.2). The organic group had an age range of 15 to 58 years with a mean age of 43.4 years (S. D. 11.4). In intelligence they ranged from a quotient of 70 to 110, the mean quotient being 94.9 (S. D. 11.7). Both sexes were represented but not in equal numbers since there were 63.2% females in the psychoneurotic group and 30.8% in the organic group.

Results

An analysis of 6 variables was made in comparing the performance of the two groups on the colored drawing test. The variable which seems most obvious and which can be most easily examined is the content of the drawings. In comparing the content of the spontaneous drawings, it will be seen from Table 1 that 94.1% of the patients with cerebral pathology produced horizontal or vertical colored strips or single-colored, undifferentiated wholes. On the other hand, only 19.1% of the patients with psychoneurosis made productions of this kind. The majority of psychoneurotics produced blended designs of various colors, objects, scenes, diagonal or wavy designs, abstract designs, or people, here designated as content of the Type 1 class.[2]

[2]It is interesting that with two patients who were not included in the present sample because of an equivocal discharge diagnosis; i.e., in both instances a strong possibility of a lesion without definite proof was indicated in addition to psychoneurosis, the patterns of the drawings were also equivocal. Both productions lacked the form of a definite design and the irregular colored areas, although not continuous strips, seemed to approach a striplike pattern. Both patients had been thought to have cerebral involvements on the basis of psychometrics alone.

TABLE 1. A COMPARISON OF THE CONTENT OF COLORED PRODUCTIONS OF A GROUP OF PATIENTS WITH PSYCHONEUROSIS AND WITH CEREBRAL ABNORMALITY

Diagnostic group	Content									
	Type 1*							Type 2		
	Blended designs (%)	Objects (%)	Scenes (%)	Diagonal or wavy designs (%)	Abstract designs (%)	People (%)	Total of Type 1 content (%)	Horizontal or vertical strips (%)	Undifferentiated one color wholes (%)	Total of Type 2 content (%)
Neurotics (N=21)	4.8	38.9	4.8	14.3	14.3	4.8	81.9	14.3	4.8	19.1
Organics (N=17)	—	5.9	—	—	—	—	5.9	58.8	35.3	94.1

*A considerable number of drawings of the psychoneurotic group could have been classified in more than one "Type 1" category, but in order to simplify the analysis each production was placed in only one such category.

The differences between the two groups in average time consumed in drawing on the test as a whole as well as for each color used were computed. Although these trends are not statistically significant ($p = .2$ in each case), the neurotic patient tends to spend a greater amount of time on the task as a whole (average time 165.2 seconds) than does the patient with a cerebral involvement (average time 135.5 seconds). The organic patient, however, tends to consume a longer period of time for each color which he uses (mean time of 83.0 compared to 55.3 seconds). These findings seem to suggest that the neurotic patient generally selects a greater number of chalks of different colors than patients with a cerebral pathology. The Chi-square analysis in Table 2 illustrates that the two groups differ significantly in the total number of chalks of different colors used.

TABLE 2. A COMPARISON OF THE TOTAL NUMBER OF DIFFERENT COLORS USED BY 21 NEUROTIC AND 17 ORGANIC PATIENTS

Number of Colors	Frequencies		X^2	p	d. f.
	Neurotics (%)	Organics (%)			
1	4.8	35.3			
2	19.0	5.9			
3	19.0	5.9			
4	4.8	11.8			
5	52.4	41.1			
			9.2	.05	4

The fifth variable analyzed involved the *kind* of colors selected first. The results presented in Table 3 indicate that patients with pathological cerebral conditions tend to use the color blue much more frequently, and the color black much less frequently than expected on a theoretical prediction of equal frequency of usage of all five colors.

TABLE 3. THE FREQUENCY OF USAGE OF EACH OF 5 COLORS AS A FIRST AND AS A LAST CHOICE FOR A GROUP OF 21 PATIENTS WITH PSYCHONEUROSIS AND 17 PATIENTS WITH CEREBRAL PATHOLOGY

	Blue	Green	Red	Yellow	Black	X^2	p	d.f.
First choice								
Psychoneurotics								
Observed frequency	9	1	3	2	6	9.9	.05	4
Expected frequency	4.2	4.2	4.2	4.2	4.2			
Patients with cerebral abnormality								
Observed frequency	12	1	3	1	0	28.6	<.01	4
Expected frequency	3.4	3.4	3.4	3.4	3.4			
Last choice								
Psychoneurotics								
Observed frequency	3	5	4	3	6	1.4	.80	4
Expected frequency	4.2	4.2	4.2	4.2	4.2			
Patients with cerebral abnormality								
Observed frequency	6	2	4	3	2	3.4	.50	4
Expected frequency	3.4	3.4	3.4	3.4	3.4			

The neurotic patient also, as can be seen from Table 3, tends not to use all colors equally frequently as his first choice. The colors blue and black both tend to be more often selected first than expected on a theoretical basis of equal frequency.

Not only may a patient's first selection of a particular color be of special significance in his dealing with emotional material, but his last selection may also be of particular importance. Accordingly a similar analysis was done for the last choice and is summarized in Table 3. The distribution of frequencies is found not to be significantly different from chance expectations for either group and therefore this measure fails to differentiate between the two groups.

DISCUSSION

The most significant finding in this pilot study is that patients with emotional disturbance are able to organize colored materials, which are assumed to be related to an individual's emotional drives, on a more complex level than the organic is able to do. The technique used, although similar in certain respects to the Rorschach test, differs in that the patient is required to make not only selective and associative responses of a perceptual type to colored stimuli, but he is also asked to actively organize and construct something by altering the environment. As Fortier[2] suggested, in his discussion of the Mosaic Test and easel-and finger painting techniques, this involves the making of a plan and its execution through actual manipulation of objects. It is for this reason that the present technique is regarded as especially sensitive for diagnostic purposes.

The patient with cerebral pathology appears to react either with emotional diffusion, as seen in his tendency to often use one colored chalk to fill in the entire page, or with excessive stereotype and "stimulus boundness", as reflected in his tendency to make simple colored strips. As Goldstein[3] has indicated, the organic suffers from an impaired ability to abstract and is overly concrete. Thus parallel to the observed behavior in dealing with colored material goes the organic's reduction of ego control over affect and his characteristic dependency upon immediate environmental stimulation. The neurotic, on the other hand, although he may overreact to emotional situations and be driven by conflicting emotions, does not usually manifest the same fundamental inability to control affect.

Summary

Numerous devices and investigations of the validity of psychological techniques aimed at detection of pathological cerebral functioning abound in the literature [1, 3, 5, 6]. The present investigation was designed to determine in what ways patients with cerebral abnormality and those with psychoneurosis differed on a spontaneous color drawing test. It was assumed that under the impact of highly emotional material, characteristic impairment of visual-motor organization would be most evident. The six variables analyzed were (1) content of drawings; (2) time consumed in actual color usage; (3) average time for each color; (4) number of colors used; (5) colors selected first; and (6) colors selected last. The most significant finding refers to the high percentage of patients with cerebral pathology who produce either a series of simple colored strips or use one color in completing the entire surface of the page. The neurotic patients, on the other hand, tend to organize their work into higher level designs. There is very little overlap between groups. Because of the relatively small number of patients used, and the fact that the groups were not strictly comparable because of the differences in age and intelligence, these preliminary results are merely regarded as trends to be cautiously interpreted and verified with larger populations.

References

1. Bender, Lauretta. A visual-motor gestalt test and its clinical use. *Research Monograph No. 3, Amer. Orthopsychiatric Assoc.*, 1938.
2. Fortier, R. H. The response to color and ego functions. *Psychol. Bull.*, 1953, *50*, 41-63.
3. Goldstein, K. and Scheerer, M. Abstract and concrete behavior. An experimental study with special tests. *Psychol. Monograph*, 1941, *53*, No. 2.
4. Hewson, Louise R. The Wechsler-Bellevue scale and the substitution test as aids in neuropsychiatric diagnosis. *J. Nerv. Ment. Dis.*, 1949, *109*, 158-183; 246 266.
5. Michal-Smith, H. The identification of pathological cerebral function through the H-T-P technique. *J. clin. Psychol.*, 1953, *9*, 293-295.
6. Piotrowski, Z. The Rorschach inkblot method in organic disturbance of the central nervous system. *J. Nerv. Ment. Dis.*, 1937, *86*, 525-537.

A PRELIMINARY STUDY TO DEVELOP A MORE DISCRIMINATING F+ RATIO

MARVIN J. FELDMAN, CAROLYN GURSSLIN, MARVIN L. KAPLAN AND NIDIA SHARLOCK

University of Buffalo

Introduction

While Rorschach [4] suggests in his original monograph that form level is related to ego control, the explicit interpretation of the F+ ratio as a measure of ego strength was made by Beck [1]. His interpretation received some experimental confirmation from a study by Williams [8] which demonstrated that F+ could be used to predict the reactions of normal subjects to a highly stressful situation. Most experienced Rorschach workers, however, have probably noted that many neurotic subjects achieve F+ ratios in the optimal range between 80-90%. Certainly these patients would not be characterized as having strong egos; hence, the general applicability of the ratio would seem to be limited. The records of neurotic subjects with high F+ ratios usually contain a preponderance of easily perceived forms, i.e., popular and vague concepts which still receive good form ratings.

Estimates of ego strength based on F+% could be more valid if it were possible to separate those clinical patients with spuriously high ratios from the total group of subjects with optimal ratios. The present study undertakes to devise such a more discriminating F+ ratio.

Form Level Ratios

At present, a number of F+ ratios are used which vary slightly and probably intercorrelate highly. Thus Beck's ratio considers only pure form responses and can be written:

$$\text{Beck F+\%} = \frac{\Sigma \text{ pure F+}}{\Sigma \text{ pure F}\pm}$$

It should be noted, however, that animal and inanimate movement responses receive a pure F score in Beck's system.

Hertz makes use of three F+ ratios called "R+", "F+ Rorschach", and "F+ primary". In this study only the R+ ratio will be used for comparative purposes. This ratio can be written:

$$\text{Hertz R+\%} = \frac{\text{all + responses with form}}{\text{all } \pm \text{ responses with form}}$$

Thus R+ includes all responses in a given record excepting formless concepts such as C, K, etc.

To better understand the construction of the new ratio, it should be noted that responses receiving "plus" form ratings in Beck's[2] or Hertz's[3] tables can be further defined into the following four classes:

1. Originals: infrequent responses listed in Hertz's tables supplemented by judgment of examiner.
2. Usuals: responses given too often to be originals but not often enough to be populars. These responses make up the majority of responses in most normal records.
3. Populars: very frequent responses as listed in Hertz's tables.
4. Vagues: responses which inherently have indefinite form are scored as vagues, e.g., clouds, smoke, indefinite maps, and x-rays. Some concepts which usually have definite form may be scored as vagues when they are not clearly delineated, e.g., a human figure seen indistinctly in the distance. Conversely some concepts which usually have indefinite form may be clearly articulated and would not be scored as vagues, e.g., a cloud which takes a human or animal form.

The numerator of Hertz's R+ ratio includes all four types of "plus" responses. The numerator of the proposed new ratio eliminates the latter two categories but uses the same denominator as in the R+ ratio. The new ratio would be written:

$$\text{Feldman's F+\%} = \frac{\text{all + form originals and usuals}}{\text{all } \pm \text{ responses with form}}$$

The writers believe that there will be fewer "false negatives" using this ratio than is true of older ratios, i.e., fewer neurotic subjects will obtain ratio scores within optimal limits. It should be apparent that the new ratio is computed in such a way that it is quite possible to obtain a 100% Beck F+ or Hertz R+ score and still be close to zero on the new ratio. This lack of interdependence between the new and old ratios enhances the possibility that the new ratio will yield results different from the older ratios although it doesn't guarantee that such will be the case.

Because the present study is limited in scope, it cannot clearly demonstrate the superiority of the new ratio over the older ones. The aim of this study, rather, is to demonstrate that the new ratio yields different results with clinical subjects than do either of the other two ratios. This finding would seem to be a necessary although not sufficient condition if the new ratio is to be a more effective predictor of behavior reflecting ego strength than its predecessors. Thus two results anticipated with the new ratio can be stated as hypotheses:

1. The scores of the clinical group will be significantly lower than those of the normal group on the new F+ ratio. Neither of the other two ratios however will differentiate between the scores of normal and neurotic subjects.
2. The correlation between scores on the new and the older ratios will be significantly higher in normal than in neurotic subjects.

Verification of the second hypothesis is necessary if it is to have an interpretation different from the older ratios, but verification of the first hypothesis actually offers some presumptive evidence for the greater efficacy of the new ratio.

Subjects

The sample populations consisted of 53 normal and 30 psychoneurotic subjects. Although perhaps adequate for the purposes of this preliminary study, the normal group was far from an ideal sample consisting as it did of practice cases obtained by members of the senior author's Rorschach class. To purify the normal group, cases were excluded who were or had been under psychiatric care and to make the group roughly comparable to the neurotic group, a lower age limit of 19 was set. The other group was composed of cases drawn from the files of a VA mental hygiene clinic and diagnosed as psychoneurotics.[1]

The two groups were not matched as closely as might be desirable. Table 1 indicates that the two groups were quite different with respect to education and average number of Rorschach responses. It cannot be decided upon an *a priori* basis whether or not a systematic bias distorts the results and in what direction such bias might operate as a consequence of the differences in education and number of responses. However, the results of this study must be interpreted with caution until confirmed by other studies.

Table 1. A Comparison of Normal and Neurotic Groups on Age, Education, and Number of Rorschach Responses

Group	N	Average Age		Average Education		Average No. of R	
		x	s	x	s	x	s
Normal	53	31.4	9.8	14.0	4.3	31.5	14.2
Neurotic	30	36.4	13.6	10.2	2.4	24.5	3.5
	t	1.73		16.0		2.39	
	p	.08		.0001		.01	

Procedure and Results

The records of the two groups were scored for form level with the aid of the Hertz frequency tables[3]. All "plus" responses were further designated as either original, usual, popular, or vague. A simple check was made on the reliability of scoring. On a random basis, 100 responses were selected from the records of the normal subjects. Six of the ten test administrators then scored this standard series. There was absolute agreement on 78% of the scores, which is well above chance expectation and affords some evidence for believing that the reliability of the scoring of individual records was adequate for the purposes of this study.

The first analysis compared the average scores between normal and neurotic groups on the three form level ratios. Table II seems to bear out the first hypothesis since neither Beck's nor Hertz's ratios discriminate between the two groups while

[1]In his position as consultant in clinical psychology, the senior author obtained a portion of his data from The Veterans Administration Regional Office, Buffalo, New York.

TABLE 2. COMPARISON BETWEEN NORMALS AND NEUROTICS OF THE AVERAGE SCORES ON THREE F+ RATIOS

Ratio	N	$\bar{x}$	s	t	p(one tailed test)
Hertz R+					
Normal	53	82.3	10.3		
				.75	.20
Neurotic	30	85.1	10.1		
Beck F+					
Normal	53	83.3	11.7		
				1.11	.10
Neurotic	30	85.1	1.14		
Feldman F+					
Normal	53	56.8	10.8		
				3.01	.001
Neurotic	30	47.7	14.2		

the new F+ ratio discriminates in the expected direction. The present data also furnish a start on the collection of normative data for the new ratio. The optimal range of scores on older ratios is generally considered to be between 80-90%. Scores on the new ratio, however, tend to be considerably lower than scores on older ratios; thus necessitating a somewhat different frame of reference for interpretations. As a first approximation, it would seem that the optimal range for the new ratio might be between 50-65%. However, these limits should be considered as highly tentative because of the inadequacy of the present normal sample. The test of the second hypothesis was based upon the intercorrelations of the three ratios in each group. The results shown in Table 3 tend to bear out the second hypothesis. The new F+ ratio correlates higher with each of the older ratios in the normal than in the neurotic sample. The difference between "r's" is not large enough, however, to be statis-

TABLE 3. INTERCORRELATIONS OF THREE F+ RATIOS IN NORMAL AND NEUROTIC GROUPS

Groups	Feldman F+ vs. Hertz R+	Feldman F+ vs. Beck F+	Hertz R+ vs. Beck F+
Normal (N=53)	.61	.37	.84
Neurotic (N=30)	.25*	.17	.87

*Difference between r's significant at .05 level.

tically significant when considering the new ratio and Beck's F+, although the difference attains significance when comparing the correlations of the new ratio and Hertz's R+. At any rate, the low magnitude of correlation between the new ratio and each of the older ratios in the neurotic group indicates that it is possible to obtain quite different results by using the new ratio. While the new ratio might yield results comparable to R+ in normal groups, it does not seem to have much in common with Beck's F+ ratio in the normal or in the neurotic group. The high correlations between Hertz's R+ and Beck's F+ in both normal and neurotic groups indicate that both ratios measure much the same thing. It seems likely that all F+ ratios now in use would also intercorrelate highly. Thus it can be concluded that F+ ratios in current use all lead to similar estimates of ego strength, but such estimates based upon the new ratio may differ markedly with those of its predecessors particularly for neurotic subjects. However, the evidence in this study is far from conclusive in demonstrating whether the new ratio yields more accurate predictions of behavior for neurotics than is true of the older ratios.

Discussion

Since the results obtained in the present study were in accordance with hypotheses stated, the new F+ ratio seems to merit further study. A more decisive proof of the efficacy of the new ratio would be to compare its predictive power with that of older F+ ratios in experimental situations like the William's study [5] where differences in ego strength should lead to differential reactions. In such studies the new ratio might have greater predictive power for neurotic subjects, but not necessarily for normals.

When we examine the nature of this new ratio more closely in terms of Rorschach rationale, implicit assumptions seem to emerge about personality dynamics. Chiefly, we assume that stressful life situations produce more anxiety for people with rigid stereotyped defenses, i.e., individuals with flexible ego structures are not so dependent upon conventional stereotyped patterns of behavior to deal with stress in their lives. By utilizing only usuals and originals in the new F+ ratio, it is believed that the new measure will more closely reflect these functions of ego strength.

Summary

1. The purpose of this study was to devise a new F+ ratio which might be a more valid measure of ego strength for certain neurotic groups than is true of F+ ratios in current use.
2. The new ratio was based upon the percentage of usual and original responses of good form to the total number of responses excepting formless concepts such as C, K. etc.
3. When the records of a group of normals and a group of neurotics were compared, there was a significant difference between the scores of the two groups on this new ratio but no difference on Hertz's R+ or Beck's F+ ratios.
4. A correlation matrix was obtained from scores on the new F+, Hertz's r+ and Beck's F+ ratios. As predicted above, the correlations were lower for the neurotic sample than for the normal sample between the new F+ ratio and each of the other two ratios. Hertz's and Beck's ratios, however, intercorrelated very highly in both samples.
5. The present findings indicate that the new ratio is sufficiently different from its predecessors to justify using it as a complement to the older ratios.

References

1. Beck, S. J. Rorschach's F plus and the ego in treatment. *Amer. J. Orthopsychiat.*, 1948, *18*, 395-401.
2. Beck, S. J. *Rorschach's test, Vol. I. Basic Processes.* New York: Grune and Stratton, 1950.
3. Hertz, Marguerite R. *Frequency tables for scoring responses to the Rorschach inkblot test.* Cleveland: Western Reserve Univ. Press, 1951.
4. Rorschach, H. *Psychodiagnostics.* Berne, Switzerland: Hans Huber publishers, 1942.
5. Williams, M. An experimental study of intellectual control under stress and associated Rorschach factors. *J. consult. Psychol.*, 1947, *11*, 21-29.

THE EFFECTS OF RECENT PERCEPTUAL TRAINING AND EXPERIENCE ON RORSCHACH PERFORMANCE[1, 2]

IRWIN J. KNOPF

State University of Iowa[3]

PROBLEM

Many investigators would agree that the structural aspects of personality are in a sense permanent, and as a result are relatively resistant to change or modification[(1, 11)]. If stability of personality structure is to be expected, then it should follow logically that any test which purports to measure personality structure should reflect this stability in its results. Such leading exponents of the Rorschach method as Beck[(2)], Klopfer[(9)], Hertz[(7)], Halpern[(6)], and Piotrowski[(10)] have, in fact, made this claim.

Numerous studies have demonstrated that under certain conditions Rorschach scores can be significantly modified. Conditions such as induced hypnotic states, drugs, faking, motor inhibition and hyperactivity, and a variety of attitudinal and instructional "sets" are among the many which have been employed. To the extent that some of these conditions were transitory and superficial, these studies cast considerable doubt on the validity of the assumption that the Rorschach reveals basic personality structure. The recent experiments which have demonstrated that various Rorschach scoring variables can be significantly affected by means of varying the "set" of the subject, however, have relied heavily on volitional factors. While these results indicate that a subject can wilfully mask or change the presumed picture of his personality structure, conscious "sets" of this sort seem to "stack the cards" against the validity of this Rorschach assumption in that the subject is told how and in what direction he is expected to change.

This study was designed to determine the "unconscious transfer" effects of recent perceptual training and experience on certain aspects of the Rorschach performance, in an attempt to shed additional light on the validity of this Rorschach assumption, as well as to extend our knowledge of conditions which may affect the stability of Rorschach performance. Certain hypotheses were formulated in conjunction with the experimental conditions.

1. The introduction of perceptual training would alter these features of the Rorschach performance:
   a. Increase the number of Dd responses (Dd%).
   b. Increase the number of pure form responses (F%).
   c. Increase the number of animal responses (A%).

2. The introduction of perceptual experience with color would modify the Rorschach performance as follows:
   a. Increase or decrease the sum of the color responses.
   b. If such alteration was found there would not be a corresponding change in the M responses.

PROCEDURE

A total of 48 college students, 24 females and 24 males between the ages of 17 and 24 were randomly selected from a list of volunteers in the Introductory Psychology courses at Northwestern University. All of the subjects were considered "naive" in that they had not been exposed to the Rorschach Test prior to this experiment.

---

[1]This is a modified version of a dissertation submitted to the Graduate School of Northwestern University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The writer wishes to express his appreciation to Drs. T. W. Richards, C. P. Duncan, E. L. Clark, and J. W. Cotton for their valuable suggestions and counsel.

[2]Presented at the annual Midwestern Psychological Association meeting in Chicago, 1953.

[3]From the Psychopathic Hospital and the Dept. of Psychiatry, College of Medicine.

The basic design of the study involved Rorschach test-retest, spaced three weeks apart, for all subjects. The standard procedure recommended by Beck [3] for the administration and scoring of the test was employed. Immediately preceding the second Rorschach testing the subjects were exposed to one of three conditions for a period of approximately 30 minutes. After the first Rorschach administration, the 48 subjects were assigned and matched to one of three groups on the basis of their scores for R, Dd%, F%, A%, sum of C, and sum of M. There were eight females and eight males in each group, making a total of 3 groups of 16 subjects. The identity of the subjects was concealed from the Experimenter by code numbers assigned independently during the matching process. Additional precautions were included to prevent "conscious" transfer from the experimental conditions to the final Rorschach performance. The subjects were informed when initially contacted that they were to serve in two separate and different experiments. Moreover, the experimental and control conditions were administered by three assistant Experimenters[4], as well as in a different building. Each assistant served an equal number of times for each of the three conditions, so that the individual effects of the assistant Experimenters were equally divided among the three conditions.

Description of the Conditions

Experimental Condition I consisted of the presentation of a series of 14 2″ x 2″ slides, half of which were achromatic, and half of which were chromatic.[5] A fifteenth slide, which was both black and white and colored, was used in conjunction with the instructions for practice purposes. In each instance there was an apparent content of structured and conventional objects. However, hidden in each of the slides as a small detail of the larger figure was one or more animals or animal parts. Each slide was exposed for ten seconds, with a ten-second interval between exposures. Four trials of 14 exposures per trial, together with the instructions and practice slide, comprised the entire session. The order of presentation of the slides was predetermined and fixed so that it was the same for each subject and the same for each trial. The subjects were required to determine the total number of hidden figures for each slide, to identify them, and to locate them.

Experimental Condition II consisted of the individual presentation of a sound motion picture film entitled "The Nature of Color". While the film involved a vivid color experience, its contents were objective and non-personal. Following the presentation of the film, the subject was asked to write all that he could remember about the movie for the bogus purpose of evaluating the effectiveness of films for college teaching.

Control Condition III consisted of a list of 14 nonsense syllables presented in a serial order on a memory drum, with the first syllable used as the start signal. Standard two-second exposures and procedures were followed. A total of forty trials comprised the entire session.

Results

Table 1 gives the mean scores obtained from the first and the second Rorschach administrations for the experimental and control groups on each of the six Rorschach scoring variables. From this it will be noted that the mean values on the initial Rorschach performance were approximately equal for all groups with respect to these scores.

In order to control for any inter-subject variation on the initial test scores which may not have been eliminated in the matching process, analysis of co-variance was applied to the data for each scoring variable. Each analysis included the variance contributed by each of the three groups. A summary of the F values ob-

[4]The author is indebted to Misses Katherine Jones, Barbara Breslower, and to Dr. Victor E. Montgomery for serving as assistant experimenters.

[5]The writer is most grateful to his wife, Roberta, for her work in constructing these slides.

TABLE 1. RORSCHACH TEST-RETEST MEAN SCORES FOR EACH GROUP OF 16 SUBJECTS

Rorschach Variable	Group I (Perceptual)		Group II (Color)		Group III (Controls)	
	Test	Retest	Test	Retest	Test	Retest
R	24.75	32.25	26.56	26.37	24.87	26.80
F%	58.14	63.31	57.69	58.62	57.69	56.76
A%	46.47	49.80	45.36	44.20	47.35	44.68
Dd%	6.18	12.42	5.48	7.95	6.55	8.79
C	3.37	3.94	3.34	2.97	3.34	3.47
M	1.87	2.00	3.06	3.50	2.75	2.69

TABLE 2. SUMMARY OF F VALUES AND PROBABILITIES FROM ANALYSIS OF CO-VARIANCE OF THE FIRST AND SECOND RORSCHACH TESTS FOR ALL GROUPS (48 Ss)

Rorschach Variable	F Value*	Probability
R	2.99	.060
F%	1.20	
A%	2.63	.062
Dd%	2.49	.087
C	.83	
M	.75	

*With 2 and 44 degrees of freedom the F values for the .05 and .01 levels of confidence are 3.21 and 5.12 respectively.

tained and their probabilities are shown in Table 2. By means of a technique introduced by Burke [4], it was possible to estimate the actual probability values of F.

The results indicate that the three groups were not significantly different on their second Rorschach performance with respect to the sum of C and sum of M scores. The F test obtained between the means of the three groups for the F% score indicates that the groups were not significantly different with respect to this variable. The F values of the R, A%, and Dd% scores, however, were statistically reliable at the .060, .062, and .087 levels of confidence respectively. Inspection of the mean difference suggests that these F values were most influenced by and attributed to the mean increases from the first to the second tests associated with Experimental Condition I. While these findings do not meet the customary criteria for statistical significance, they may reflect important trends in terms of the effect of perceptual training on the Rorschach performance.

In an effort to elucidate these marginal findings, this Experimental condition was repeated with 11 additional subjects, five females and six males. The subjects were obtained in the same manner as in the first experiment, and wherever possible the procedure and experimental design were identical.

Mean scores for the first and second Rorschach testing and the mean differences between the two test performances are listed in Table 3. The results obtained from separate analyses of co-variance indicate no statistically reliable F value for any of the scores in question, namely: R, F%, A%, or Dd%. However, in each instance the mean difference was in the same direction as previously found and also was nearly of the same magnitude.

TABLE 3. RORSCHACH TEST-RETEST MEAN SCORES AND MEAN DIFFERENCES FOR THE REPLICATED GROUP OF 11 SUBJECTS

Rorschach Variable	Mean Scores		Mean Differences
	1st Test	2nd Test	
R	23.73	26.00	2.27
F%	60.42	69.31	8.89
A%	48.66	55.91	7.25
Dd%	6.80	11.87	5.07

DISCUSSION

The results obtained from Experimental Condition II (the film on color) indicated clearly that this experience failed to alter appreciably the sum of C or the sum of M scores on the Rorschach performance. Insofar as this condition is concerned, these findings would seem to lend support to those who believe that the "core" of the personality as indicated by the Rorschach method changes very little. On the other hand, the results obtained from Experimental Condition I (perceptual training) and its replication consistently point to an increase in the R, A%, and Dd% scores as a consequence of this training experience. It is apparent, however, that the increases in scores were small, and in order to demonstrate this effect at a statistically reliable level it would be necessary to increase the size of the sample considerably to reduce the error term in order to obtain large enough F values. At best, therefore, it could be said that while this experience has increased these Rorschach scores, the effect is small and perhaps negligible.

The results of other studies indicate clearly that "set" is an important variable affecting Rorschach performance. Under conditions of specific instructions to change, Hutt *et al*[8], and Gibby[5] found that the F+, D+, Dd, W, R, range of content, and A% scores were significantly altered, while no concomitant modification was observed in the M:C ratio and the M and C scores. However, when the subjects were instructed to perceive more M, this score significantly increased. The findings of the present study may reflect the effects of two other points on the continuum of "set". The perceptual training experience involved a less direct "set" in which the subjects were given an orientation to perceptual material. While this condition did seem to increase some scores slightly, mere exposure to a film on color did not affect the stability of the Rorschach scores studied. Thus, it would appear that there is a relationship between the strength of the "set" and the modification of Rorschach performance.

Nevertheless, at least two alternative explanations remain: (1) certain aspects of the Rorschach performance can be altered more readily than others, and (2) interaction between "set" and the stability of some Rorschach scores may be operating here. While the present data do not provide an unqualified choice of one of these alternatives, they do suggest that similar explorations of Rorschach scoring variables under a wide variety of conditions would be helpful in reducing sources of errors, and increasing the validity and predictability of Rorschach interpretations.

SUMMARY

The effects of two experimental conditions on six Rorschach scoring variables were studied. One group of 16 subjects was trained to perceive small animals or animal parts hidden in the apparent content in each of a series of 14 slides. Another 16 subjects were exposed to a non-personal film on color. A control group of 16 subjects learned a serial list of nonsense syllables. All subjects were given the Rorschach

twice, spaced three weeks apart, with the experimental conditions immediately preceding the second Rorschach testing. The results indicated that the color experience had no appreciable effect on the sum of C or sum of M scores, whereas the R, A%, and Dd% scores under the perceptual training condition very closely approached statistical significance at the .060, .062, and .087 levels of confidence respectively. The F% score for the latter condition was not significantly affected. The perceptual training condition was repeated with 11 additional subjects. While the resulting mean differences were in the same direction as previously found, separate F tests yielded values smaller than those required for the customary levels of significance. These results point to the slight and perhaps negligible effect of this perceptual experience on the Rorschach performance.

Bibliography

1. Allport, G. W. *Personality: A Psychological Interpretation.* New York. Henry Holt and Co., 1937.
2. Beck, S. J. Stability of the personality structure. *Psychol. Bull.*, 1942, 39, 512.
3. Beck, S. J. *Rorschach's Test:* I. *Basic Processes.* (2nd Ed.) New York: Grune and Stratton, Inc. 1950.
4. Burke, C. J. Computation of the level of significance in the F-test. *Psychol. Bull.*, 1951, *48*, 392-397.
5. Gibby, R. G. The stability of certain Rorschach variables under conditions of experimentally induced sets: I. The intellectual variables. *J. Proj. Techniques*, 1951, *15*, 3-26.
6. Halpern, R. Rorschach interpretation of the personality structure of schizophrenics who benefit from insulin therapy. *Psychiat. Quart.*, 1940, *14*, 826-833.
7. Hertz, M. R. The validity of the Rorschach method. *Amer. J. Orthopsychiat.*, 1941, *11*, 512-520.
8. Hutt, M. L., Gibby, R. G., Milton, E. O., and Potthorst, K. The effect of varied experimental "sets" upon Rorschach test performance. *J. Proj. Techniques*, 1950, *14*, 181-187.
9. Klopfer, B. Personality aspects revealed by the Rorschach method. *Rorschach Res. Exch.*, 1940, *4*, 26-29.
10. Piotrowski, Z. A. On the Rorschach method of personality analysis. *Psychiat. Quart.*, 1942, *16*, 480-490.
11. White, R. W. *The Abnormal Personality.* New York: Ronald Press Co., 1948.

CONTRIBUTIONS TO A THEORY FOR SELECTING PSYCHOTHERAPY PATIENTS

FRANK AULD, JR. AND JEROME K. MYERS[1]

Yale University

Introduction

Selecting suitable patients for psychotherapy is an important task, in which the therapist welcomes help from any quarter [17, p. 39]. In his decision to accept or reject a patient he is quite often influenced by the results of psychological tests. There is no general agreement, however, on what test results ought to persuade a therapist to accept a patient, and what results ought to induce him to reject a patient. Obviously, then, research on the use of psychological tests in selection of patients is much needed.

Research on the validity of tests as guides to selection of patients can follow either of two broad lines. The first approach may be called *empirical.* The investigator who takes this approach chooses some criterion of suitability of the patient (e. g. the therapist's judgment of the success of treatment or the number of interviews the patient stayed in therapy) and finds out what tests or test items are cor-

[1]Dr. Leslie Schaffer, formerly of the Yale Department of Psychiatry and now serving in the Army, collaborated with Myers in the collection of the data presented in this paper. Undoubtedly his ideas, as well as the facts he helped to gather, have shaped our thinking on the matters discussed in this paper. We are glad to acknowledge Dr. Schaffer's important contribution. The investigation was partly supported by a research grant from the U. S. Public Health Service.

related with this criterion measure. Any test or test item that is correlated with the criterion is thereafter used as a predictor, i.e., is used to select patients. The investigator is usually not interested in how or why the test predicts; he is satisfied if it seems to predict.

The other approach might be called *rational,* because the investigator taking this approach attempts to understand the *how* and *why,* then tries to make predictions. He decides first what traits make a patient suitable for psychotherapy (in deciding this, he is probably guided by clinical experience, since no systematic studies exist), then he tries to measure these traits with reliable and valid instruments. The investigator taking this approach reaches the stage of application later, since he has to know what his tests are measuring before he can use them. But (we shall argue) his results are more likely to stand the test of time.

Examples of the empirical approach are not hard to find. Kotkov and Meadow [9] and Rogers, Knauss, and Hammond [15] did empirically-minded studies, using as the criterion length of therapy; and Harris and Christiansen [8] and Rogers and Hammond [14] did similar studies, with success of therapy as the criterion. No example of a study in which the rational approach was used is known to the authors.

Difficulties of the Empirical Approach

The investigator who takes the empirical approach is beset by many difficulties. If he tries out a large number of test items, by chance he will discover some of them to be correlated with his criterion. This makes cross-validation studies necessary. Furthermore, the investigator cannot be sure that the correlations between test items and criterion hold generally true for all prospective patients, or whether the correlations he discovered are only true for the particular type of patient he happened to study. This makes validity-generalization studies necessary [11]. But even more important, the investigator doesn't understand very well why his test items are correlated with his criterion. Of course, he can try to explain the correlations after he gets them. Many ingenious *ad hoc* explanations have been invented! But often the investigator gives the wrong reasons instead of the right ones.

Not knowing why he could make the predictions is a serious handicap to the investigator. Because of his ignorance, he can't tell in which samples his prediction equations will work and in which ones they will not. Furthermore, he is stuck with the test items he first tried out, even if these happen to be laborious and awkward measures of the trait that makes prediction possible. For instance, if it should turn out that the total number of responses to the Rorschach predicts continuance in psychotherapy *only* because this item is a measure of verbal intelligence, it is obvious that measuring intelligence in this way would be inefficient. The verbal scale of the Wechsler-Bellevue would be easier to administer and would give better predictions [3].

First Steps Toward a Theory for Predicting Length of Psychotherapy

Since we believe that the rational approach has many advantages over a purely empirical one, we shall here try to make a start on constructing a theory for predicting length of psychotherapy. In this paper we wish to discuss only one characteristic of the patient, his social position. We realize, of course, that a comprehensive theory for selecting patients will have to embrace a large number of aspects of the patient's background, personality, and current life circumstances.

Dollard found that some middle-class patients come to therapy almost ready to jump on the couch and begin free association, without need for a word of instruction![2] Upper-middle-class patients generally have heard of psychoanalysis and psychotherapy; they know what a psychiatrist does, that he helps patients by talking with them. Lower-class patients are apparently less well-informed. Their experiences with doctors are with physicians who have treated physical complaints.

[2]Personal communication from Dr. John Dollard.

As Freud said [7, pp. 345-346], belief in the efficacy of psychotherapy is probably of little importance; nevertheless, misinformation or wrong expectations could be an obstacle to the therapeutic work. Colby [5] has noticed that the middle-class patient is more likely to be "psychologically minded." He feels that this greater receptivity to therapy is a definite advantage.

The social position of the patient not only influences what he knows and what he expects, but also seems to determine to some extent what rewards he will get out of therapy.[3] Patients are rewarded in many ways by psychotherapy: in the ideal case, they are rewarded when they learn and try out new, more adaptive habits; they are also rewarded by the therapist's attention, by his understanding, and by his empathic interest in them. To get the rewards of more adaptive habits, life conditions must offer rewards for adaptive behavior. Perhaps in some cases the lower-class patient's life has little to offer to reinforce a change in behavior; or, even if changed behavior could be rewarded, the middle-class therapist may be unfamiliar with the conditions of lower-class life and therefore unable to act as a dependable guide to the patient. To get the rewards of the therapist's understanding and interest, the patient must make himself understood. He finds it harder to do this if he is a lower-class patient. The therapist doesn't understand his language or the conditions of his life. Finally, to be rewarded by the therapist the patient must be able to understand what the therapist is saying. This is less often possible when the patient is lower-class and the therapist middle-class. Moreover, when we turn to consider the punishments in the situation we see that the fee (even if it is a small fee) is more punishing to the lower-class patient than to the middle-class one.

Not only does the lower-class patient get less reward and more punishment from psychotherapy, but also his therapist gets less reward from treating him. The middle-class therapist is unfamiliar with the conditions of life of the lower-class patient. Therefore he finds it harder to be sympathetic and genuinely interested. He finds it harder to have empathic reactions to what the patient tells him. He confronts a more difficult task in instructing the lower-class patient, who, because of his different class-ways, may seem stupid.

For all these reasons, therefore, we should expect that lower-class patients would do less well in psychotherapy than middle-class patients. We should expect that their cases would be judged as less successful and that they would quit or be dismissed more quickly.

An Empirical Study of the Relation of Social Position to Length of Treatment

With this background of theory, we proceeded to find out whether our expectations would be borne out by data. Schaffer and Myers [12, 16] made a comprehensive investigation of the workings of an outpatient psychiatric clinic. Their investigation provides us with data concerning the relation of social position to length of psychotherapy.

Schaffer and Myers measured patients' social position by Hollingshead's Index of Social Position [see 13]. This Index is computed from scores on three variables: occupation, education, and ecological area of residence. The Index scores permit the assignment of each patient to one of five social levels. *Class I* comprises families of wealth, education, and top social prestige. *Class II* consists of families in which the adults for the most part hold college degrees and in which the fathers have professional or high-level managerial occupations. *Class III* includes proprietors, white-collar workers, and skilled workers; they are mostly high-school graduates. *Class IV* consists largely of semiskilled workers and laborers, with less than high-school education. *Class V* includes unskilled and semiskilled workers, who have a grade-school education or less, and who live in the poorest areas of the community.

---

[3] Our analysis of the interaction between patient and therapist parallels the analysis of teacher-pupil interaction so well done by Davis and Dollard [6, pp. 279-290].

The dependence of the number of interviews on social position is shown in Table 1. The mean number of interviews for the Class II and Class III patients was 24.7; the mean number for the Class IV and Class V patients was 11.0. To get a measure of the degree of the relationship between length of treatment and social position, we computed a biserial correlation. The biserial r between class (the dichotomized variable) and number of interviews is +.403, which is significant at the .01 level.

Discussion

The results of the study of the outpatient clinic are consistent with the hypothesis that social position of the patient influences his response to psychotherapy. These results strengthen our belief that in the interactive process between patient and therapist, there are more rewards for both patient and therapist when the patient belongs to the middle class. Another factor, however, probably also plays a part. Schaffer and Myers[12, 16] found that the Class II patients were likely to be assigned to the senior staff, the Class III and IV patients to residents, the Class V patients to medical students (patients assigned to medical students were not included in Table 1, however, because students worked in the clinic for clerkship periods of only four weeks and this arrangement set an artificial limit on the length of therapy of their patients). Because of this, it can be seen that the middle-class patient is likely to get a better-trained therapist. It seems reasonable to suppose that those patients who get more experienced, more skillful therapists will stay in therapy longer.[4]

Table 1. Length of Treatment and Social Class

Number of interviews	Per cent of patients	
	Classes II and III	Classes IV and V
1 to 7	25.0	44.0
8 to 19	32.5	36.0
20 and over	42.5	20.0

Note.—All patients reported on in this table were treated by residents or senior staff. There are 40 patients in Classes II and III and 25 in Classes IV and V.

Since length of therapy is related to the social position of the patient, one could predict from the patient's social position how long he will stay in therapy. A word of caution is in order, however; we are not sure that the relationship between social position and length of treatment would hold in other clinics. This must be determined by further studies.

Any measure of social position might give some prediction of how long patients will stay in psychotherapy. For example, some Rorschach scores may measure social position in a rough way, as Auld[1, 2] found. A study by Beck *et al*[4] produced similar findings; executives, skilled workers, and semiskilled workers gave an average of 35 responses to the Rorschach, but unskilled workers had an average of only 25. The correlation of test scores with social position would make the scores rough measures of social position; and this might permit one to predict length of therapy by means of the test scores. However, if social position of the patient is the only variable that makes prediction of length of therapy possible by these scores, we believe it would be wiser to use a precise measure of social position (like Hollingshead's Index) rather than a rough and awkward measure.

---

[4]Our explanation here is similar to one given by Miles *et al.*[10] for a relationship between IQ and success of therapy.

Summary and Conclusions

Research on the validity of tests for selecting psychotherapy patients can follow two approaches: empirical and rational. The purely empirical approach is beset by many difficulties. It is argued that the rational approach is likely to be more fruitful.

In this paper we have tried to make a start on constructing a theory for predicting length of psychotherapy. We showed why it should be expected that middle-class patients should stay in therapy longer. The experience of one psychiatric outpatient clinic was examined. It was found that in this clinic middle-class patients stayed in therapy longer. This finding was interpreted as being consistent with the hypothesis concerning effect of social position on the patient's response to psychotherapy.

We do not believe that our findings justify selecting middle-class patients and rejecting lower-class patients. (As a matter of fact, Schaffer and Myers found that the clinic unwittingly was doing just that.) Instead, we feel that these facts underline the need for therapists and administrators to understand better the social background of their patients. Perhaps, too, therapists should consider changes in technique to meet the needs of lower-class patients, as Freud suggested many years ago (7, pp. 401-402).

References

1. Auld, F. Influence of social class on personality test responses. *Psychol. Bull.*, 1952, 49, 318-332.
2. Auld, F. The influence of social class on tests of personality. *Drew Univ. Stud.*, 1952, No. 5.
3. Auld, F. and Eron, L. D. The use of Rorschach scores to predict whether patients will continue psychotherapy. *J. consult. Psychol.*, 1953, 17, 104-109.
4. Beck, S. J., Rabin, A. I., Thiesen, W. G., Molish, H. and Thetford, W. N. The normal personality as projected by the Rorschach test. *J. Psychol.*, 1950, 30, 241-298.
5. Colby, K. M. *A primer for psychotherapists.* New York: Ronald, 1951.
6. Davis, A. and Dollard, J. *Children of bondage.* Washington, D. C.: American Council on Education, 1940.
7. Freud, S. *Collected papers.* Vol. II. London: Hogarth Press, 1924.
8. Harris, R. E. and Christiansen, C. Prediction of response to brief psychotherapy. *J. Psychol.*, 1946, 21, 269-284.
9. Kotkov, B. and Meadow, A. Rorschach criteria for predicting continuation in individual psychotherapy. *J. consult. Psychol.*, 1953, 17, 16-20.
10. Miles, H. H. W., Barrabee, E. L., and Finesinger, J. E. Evaluation of psychotherapy. *Psychosom. Med.*, 1951, 13, 83-105.
11. Mosier, C. I. Problems and designs of cross-validation. *Educ. psychol. Measmt*, 1951, 11, 5-11.
12. Myers, J. K. and Schaffer, L. Social stratification and psychiatric practice: a study of an outpatient clinic. *Amer. sociol. Rev.*, 1954, 19, in press.
13. Redlich, F. C., Hollingshead, A. B., Roberts, B. H., Robinson, H. A., Freedman, L. Z. and Myers, J. K. Social structure and psychiatric disorders. *Amer. J. Psychiat.*, 1953, 109, 729-734.
14. Rogers, L. S. and Hammond, K. R. Prediction of the results of therapy by means of the Rorschach test. *J. consult. Psychol.*, 1953, 17, 8-15.
15. Rogers, L. S., Knauss, J. and Hammond, K. R. Predicting continuation in therapy by means of the Rorschach test. *J. consult. Psychol.*, 1951, 15, 45-57.
16. Schaffer, L. and Myers, J. K. Psychotherapy and social stratifications: an empirical study of practice in a psychiatric outpatient clinic. *Psychiatry*, in press.
17. Thorne, F. C. *Principles of personality counseling.* Brandon, Vt.: Journal of Clinical Psychology, 1950.

A STUDY OF TWELVE-YEAR-OLD RECIDIVISTS

WILLIAM W. WATTENBERG AND FRANK QUIROZ

Wayne University

PROBLEM

The present study is a follow-up to an investigation [8] of factors associated with recidivism among eleven-year-old boys in trouble with the police in Detroit. In that study it had been found that young boys with more than one police contact tended to have poor school records, low mental ability, membership in unruly gangs, and a reputation for mischief. The connection between repeating and the usually reported variables of poor family conditions and low socio-economic status was present but was not strong enough to produce tabulations with statistical reliability sufficient to warrant rejection of the null hypothesis at the two per cent level of confidence.

It was decided to repeat the study among twelve-year-olds because evidence in the literature indicated that reported delinquency showed an upsurge in frequency during that year of life. This was especially evident in research reports based on British experience. In describing a study made in 1948 of young people in Birmingham (England), Reed [6] stated that the peak age for juveniles referred to the courts was twelve. In an earlier study with a larger group, Carr-Saunders and his coworkers [2] had found a peak at thirteen.

In American experience, the modal age comes later if the total of boys in trouble with the police or courts in any given year of life is taken as the index. It was fifteen or sixteen according to Banay's analysis [1] of U. S. Children's Bureau statistics. The Manns [5] found that fifteen-year-olds furnished the largest proportion of delinquents. However, if age at first court appearance is taken as the index, the early upsurge becomes evident. The Gluecks [3] reported that various measures of central tendency fell between the twelfth and thirteenth birthdays.

This situation appears to be the culmination of a long process rather than the result of any sudden change in juvenile behavior patterns. Writing from the vantage point of work in a school system, Kvaraceus [4] noted that in most instances youngsters were referred to the courts only after the police had given them several opportunities to "mend their ways." Presumably, the police encounter quite a few young boys who are reported for various offenses but who merit lenient treatment. If released to their parents with a warning, such boys do not get in trouble again. By contrast, the group which finally appears in court do not respond to this type of handling. Irritated or discouraged by the persistence of misconduct and neighborhood complaints in their cases, a police officer finally decides they are beyond his resources and invokes formal court procedures.

It is important to recognize that although in the cases of boys declared delinquent by courts there is almost always a history of early misbehavior, this does not imply that all boys who get in trouble at a tender age go on to become seriously delinquent. Detroit Police Youth Bureau records reveal that more than three-quarters of the boys ten or eleven years old at the time of their first police contact have had no complaints registered against them in a subsequent year. This poses a very practical question: Which items of information, alone or in combination, discriminate between those youngsters whose early misconduct is predictive of serious trouble and those who can wisely be released to the resources of their families and neighborhoods? Since few communities have adequate facilities to deal with all children who need special handling, it is important to screen out those boys whose development would lead to a happy outcome. Then, it would be possible to use the special facilities for those youngsters in whose lives special treatment might be crucial.

For purposes of prognosis, among the eleven-year-olds previously studied [8], factors linked to school adjustment and to gang activities had proved of greater significance than those involving family situations and neighborhood conditions. The

present study was undertaken to find out whether or not this pattern was equally true for twelve-year-olds. It was realized, of course, that the mere addition of one year of chronological age would hardly produce any drastic reversals. Continued school maladjustment and membership in unruly gangs would, if anything, become more potent influences as a boy aged. However, in view of the increase in delinquency at this age level, it was suspected that the twelve-year-olds with repeated police contacts would be more similar to older delinquents in the sense that family and neighborhood would rise in significance for them. At any rate, this was the hypothesis we set out to test.

Procedure

With assistance from the Youth Bureau of the Detroit Police Department* the police records of all boys "interviewed on complaint" in 1950 were obtained. Those who were twelve years old at the time of contact, some 460 in all, were selected for study. The records indicated whether the boy had been involved in previous complaints. On this basis, the group was divided into 130 repeaters and 330 non-repeaters.

In connection with each contact, special Youth Bureau officers had made an investigation which included interviews with the boy, visits to his home, and exchange of information with the school and with other social agencies known to be interested. The information thus secured was recorded on a "history sheet," coded, and punched on IBM cards. For 1950, these contained fifty items relating to the boy, his school, his family, and his neighborhood. For each such item, a table was prepared comparing the repeaters and the non-repeaters. All tables were submitted to the chi-square test to determine the statistical significance with which the null hypothesis could be rejected.

Findings

Of the fifty tables, twelve showed relationships significant at the one per cent level; four, at the two per cent level. These are listed in Table 1. The number of statistically significant items was sixteen times chance expectation. As in the study of eleven-year-olds, school adjustment and a very rough indication of mental ability were related to repeating. At the one per cent level of confidence, more of the repeaters: (1) were rated of "below average mentality," (2) expressed unfavorable attitudes toward school, (3) were receiving poor marks, and (4) were out of their regular grades.

The relationship between gang activity and repeating likewise remained significant. At the two per cent level of confidence, more of the repeaters belonged to definitely organized gangs. More interesting was the officers' classification of the group's activities. On the basis of their knowledge, the policemen considered that the majority of the non-repeaters belonged to orderly groups. By contrast, most repeaters belonged to gangs regarded as "rambunctious;" a significant minority went in for organized theft.

As predicted by our hypothesis, family factors were also significantly related to repeating. At the one per cent level of confidence, more repeaters expressed unfavorable attitudes toward their homes and fewer of their parents took a forgiving or understanding attitude toward them. At the two per cent level of confidence, more of the repeaters came from homes in which the parents seldom or never participated in the boy's activities. The officers found that almost all parents of non-repeaters were making an effort to understand or back up their sons.

Before we can accept at face value the finding that family relationships were statistically significant for the twelve-year-olds but not for the eleven-year-olds we must take into account the fact that the population of twelve-year-olds in trouble

*The authors wish to express their gratitude to Senior Inspector Sanford Shoults, Inspector Ralph Baker, and Lieutenant Francis Davey for their aid not only in making records available but for giving generously of their time in securing the basic tabluations upon which this research is based.

TABLE 1. ITEMS LINKED WITH REPEATING

Subject of Tabulation	Item Positively Associated with Repeating	Level of Confidence (P)
Grade in school	Retarded or in special class	.01
Attitude toward school	Unfavorable	.01
School marks	Poor	.01
Rating of intelligence	Below average	.01
Gang membership	Member	.02
Gang activity	Disorderly; organized theft	.01
Attitude toward home	Unfavorable	.01
Attitude of parents	Punitive, rejecting	.01
Parents' participation in boys' activities	Seldom; never	.02
Condition of dwelling unit	Substandard; needs repairs	.02
Rating of neighborhood	Slum	.01
Family owns a car	Yes	.01
Appearance of boy	Slovenly	.01
"Type" of boy	"Peck's Bad Boy"	.01
Religion	Non-Jewish	.02
Attitude toward officer	Evasive	.01

with the police was larger. Our studies included 460 twelve-year-olds but only 334 eleven-year-olds. The chi-square test is peculiarly susceptible to the size of the group being studied. Therefore, it was necessary to go behind the simple test of significance to make sure that the apparent increase was not an artifact of the size of population. This was done by comparing the percentages for the categories in each subgroup. On this basis, no important shift was found for the tables having to do with the parents' participation in the boys' activities. Surviving this second test was the increase in significance for the boys' open expression of dislike for their homes. Among the eleven-year-olds, 4.0 per cent of the repeaters and 3.4 per cent of the non-repeaters had voiced such attitudes. Among the twelve-year-olds, 6.1 per cent of the repeaters but only 1.8 per cent of the non-repeaters expressed hostile attitudes.

On the matter of socio-economic indices, the results were more clearcut. Three such variables showed a statistically significant relationship to repeating. At the two per cent level of confidence, repeaters tended to come more often from dwelling units rated as needing repairs or being substandard. At the one per cent level of confidence, they tended to come from neighborhoods rated by police officers as "slums" and from families not possessing automobiles.

Here, again, it was necessary to go behind the simple calculation of significance to determine whether or not the apparent rise to reliability of these factors was an artifact of the increase in size of the population concerned. Comparing percentages, it appeared that the condition of housing presented a dubious case. However, on the other two items a genuine shift appeared to have taken place. The rating of "slum" had been given to the neighborhoods of 13.1 per cent of the eleven-year-old repeaters

and 14.3 per cent of the non-repeaters; for the twelve-year-olds, to 24.6 per cent of the repeaters as against 14.2 per cent of the non-repeaters. Families not possessing cars among the eleven-year-olds involved 59.6 per cent of the repeaters and 48.9 per cent of the non-repeaters as compared with 67.8 per cent of the twelve-year-old repeaters and 51.0 per cent of the non-repeaters.

Four other tabulations, not previously reported, also rose to statistical significance. At the one per cent level of confidence, the twelve-year-old repeaters were found to have been rated more frequently by the Youth Bureau officers as (1) slovenly in appearance, (2) being like "Peck's Bad Boy," and (3) "evasive" in their behavior toward the interviewing officer. At the two per cent level of confidence, Jewish boys were absent from the ranks of the repeaters. A comparison of percentages showed that, with the exception of the slovenly appearance, these were not new relationships but owed their newly acquired statistical significance to the increase in size of population.

Discussion

The findings tended to uphold the original hypothesis. School adjustment and gang activities retained the statistically significant relationship to repeating which had prevailed among the eleven-year-olds. In line with the prediction, one index of poor family relationship and two items reflecting low socio-economic status joined the list of significant items in a manner which reflected an alteration in relationships rather than a spurious statistical effect. In addition, the new statistical significance of ratings of personal appearance gave additional credence to the view that the twelve-year-old repeaters did conform more closely than had the eleven-year-olds to popular stereotypes of "the juvenile delinquent."

There is no surprise, and little point to discussion, in the fact that school adjustment and the nature of the boys' gang activities should be related to repeating. In almost all studies, school retardation has been found more prevalent among delinquents of all ages. In an earlier study Wattenberg[7] had found that juvenile arrest statistics showed an unseasonal decline in the first month of summer vacation periods. There seems little question but that school maladjustment is at least a precipitating factor in the delinquency of some youngsters. Similarly, boys who belong to gangs which engage in anti-social conduct are prone to continue getting in trouble along with their companions.

Turning to the stronger effect of bad parent-child relationships among twelve-year-olds, it should be pointed out that the new element in the picture seems to be a more openly hostile attitude on the part of the boy. There is some question as to whether this is a cause or an effect of repeated police contacts. It may well be that the pressure within the family resulting from repeated official action may embitter the youngster. Where the home situation was poor, however, the boy might turn with greater intensity to get satisfactions with his gang. This could well lead to his making a more open emotional break with his parents. It would also serve as a practical guarantee that he would continue getting in trouble. Thus, although parent-child relationships might have relatively weak influence in preventing initial experimentation with adult-challenging behavior they could be critical in determining its continuation over a period of even one year. We should expect to find among twelve-year-old repeaters a larger residual group of youngsters expressing a dislike of their homes.

This may provide partial explanations as to why the police finally refer the boys to courts and why bad family situations are so common among youngsters finally adjudged delinquent by court action. Boys reaching the courts are not a cross-section of all offenders. Rather, they are those left over when the less forceful elements of family "education" have patently failed and police have given up hope.

The rise in potency of socio-economic factors can be explained in much less complex a manner. In their initial efforts at independence, youngsters in almost any neighborhood would do some things which could bring complaints to the police. In

the areas with high delinquency rates, they have more models to copy. This could also explain why they seemed to conform to stereotyped conceptions of "tough kids." Their sloppy appearance could well reflect efforts to ape an older group of neighborhood "heroes." By twelve years of age, apparently, their induction would be in full swing. Recognizing this, the police could be expected to start filing official petitions against them.

To test further the possibilities set forth in this section, a series of longitudinal studies is required.

Summary

A group of 460 twelve-year-old boys contacted in 1950 by Detroit police was studied to find out what factors distinguished recidivists from non-recidivists. As has been found in a previous study of eleven-year-olds, school maladjustment and the nature of gang activities were significantly related to repeating. Unlike the younger group, the twelve-year-old repeaters came from a statistically significant larger proportion of homes with poor parent-child relationships and low socio-economic standing. They also bore greater similarity to stereotypes of "the juvenile delinquent" in appearance.

References

1. Banay, R. S., *Youth in Despair.* New York: Coward-McCann, 1948.
2. Carr-Saunders, A. M., Mannheim, H., and Rhodes, E. C. *Young Offenders.* New York: Macmillan, 1944.
3. Glueck, Sheldon and Glueck, Eleanor T. *Unraveling Juvenile Delinquency.* New York: Commonwealth Fund, 1950.
4. Kvaraceus, W. C., *Juvenile Delinquency and the School.* Yonkers: World Book Co., 1945.
5. Mann, Cecil W. and Helene, P., Age and intelligence of a group of juvenile delinquents. *J. Abn. Soc. Psychol.*, 1939, 34, 351-360.
6. Reed, B. H., *Eighty Thousand Adolescents.* London: George Allen and Unwin, 1950.
7. Wattenberg, W. W., Delinquency during the summer months. *J. Educ. Res.*, 1948, 42, 253-267.
8. Wattenberg, W. W., Factors associated with repeating among preadolescent "delinquents." *J. Genet. Psychol.* (Accepted for publication.)

THE USE OF DRAWINGS OF THE HUMAN FIGURE AS AN ADJUNCT IN PSYCHOTHERAPY

FRANCIS W. KING

Dartmouth College

Introduction

Despite the growing literature on the evaluative and predictive roles of projective techniques in relation to various kinds of treatment, few references exist on the direct use of projective techniques within the therapeutic situation.[1] The therapeutic use of such techniques is an almost unexplored area (10). It is primarily the Thematic Apperception Test, and its variants, that has been employed directly by clinicians in the therapeutic process (1-9). The major purpose of this brief report is to illustrate by a single case the usefulness of employing human figure drawings as an adjunct in the course of psychotherapy. It seems to the writer that this paper may also be considered as a minor contribution to the diagnostic value of at least one characteristic of human figure drawings.

The Setting

A twenty-year-old male began an extended series of psychotherapeutic interviews in the middle of his junior year in college. There were none of the frequent

[1]This, of course, does not pertain to the use of play, art, and drama; these techniques may be considered, according to their use, as projective techniques or therapeutic techniques.

academic complications for he was a brilliant and efficient student who was elected to Phi Beta Kappa during this same year. If diagnostic labels were to be employed, the term obsessive-compulsive could not possibly be omitted, for this highly effective young man scheduled his activities on college vacations as rigorously as he planned his college course work—and he abided by his schedules! "Rational," "logical," and "intellectual" were his favorite words—and defenses. Additionally, it should be mentioned that his history was marked by a number of homosexual relationships.

A series of nearly thirty interviews throughout a semester was interrupted for sixteen weeks by the college summer vacation. Until this lengthy interruption occurred, therapy had been progressing in a painfully slow manner, but nonetheless progressing. Dynamic material of some depth had occupied a considerable portion of the therapeutic hours and some significant, if small, changes took place in attitude and behavior. Homosexual activity had ceased and there was some loosening up here and there of his many rigidities.

However, when the interviews were resumed in the fall, weeks passed without any apparent therapeutic progress. Detailed accounts of the vicissitudes of a continuing friendship made during the summer occupied seemingly endless hours of interviews. Finally, in the hope of redirecting the patient's attention to less externalized matters, the therapist introduced the draw-a-person technique into the latter part of one of the interviews.

Interview Material

Figures 1 and 2 reproduce the drawings obtained during this hour.[2]

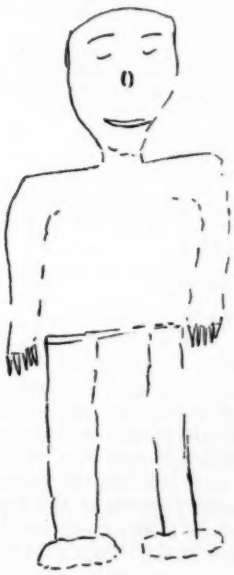


Figure 1

Figure 2

The following verbatim transcription presents the patient's comments while executing the drawings and his most revealing subsequent interpretations and associations.

[2]The original drawings, only faintly sketched for the most part, have been slightly darkened for better reproduction.

Therapist. Would you draw a picture of a human being?

Patient. (*Laughs*) Well, in elementary school and in high school, there were two subjects with which I used to come home with E's on my report card—penmanship and drawing. That was something I could never do.

T. I'll not ask you to write, just to draw a picture.

P. I haven't done this since . . . my niece could do a better job (*erases*) (*laughs*). I hope no one ever sees this. Let's see, this is supposed to be a profile. I think it would be easier if I drew a (*erases*) . . .

T. Turn the sheet over if you like.

P. (*Turns paper over and starts new drawing*) In elementary school or in high school—I guess it was in elementary school—the fellow who sat next to me used to do my drawing and I used to do his math. I think this is the most difficult thing . . . (*erases*) . . . fingers, modern art. (*Last phrase unclear*)

T. Pardon?

P. I said, modern art, and that is supposed to be a human being.

T. Fine. Now I want you to draw a picture of the opposite sex.

P. Is that supposed to have a sex? (*points to Figure 1*) I guess that would be a man. (*He starts to draw a profile again with much erasure*) I'm trying to do a profile again (*turns sheet over*) . . . (*laughs*) . . . (*erases*) A little out of proportion. Looks the same. This is, as I said a little while ago, not my field. I . . . find that this was something I could never do—was drawing anything that looked like anything; it was something I always wanted to do. My father, on the other hand, draws very well—not very well—at least he's able to do it. And drawing seemed to me to be a . . . skill of some sort which was similar to athletics possibly in some ways.

T. What sort of a person is that one? (*pointing to Figure 1*)

P. That looks like something out of Buck Rogers—or a robot. Certainly it looks well built, broad shouldered, muscular, I guess. It looks like something a two-year-old drew. It always seemed to me that something you could do with your hands, a skill, was more worthwhile than something you thought about or came to through thinking. When I was young, I guess, oh, five or six or so, I used to be a wonderful mechanic and very handy with tools. I used to take everything in the house apart, and frequently not put it together again. I used to know the uses of all the tools. My mother tells me that my father was quite sure I was going to be an engineer, and very soon that interest and skill disappeared, although some of the interest was still there. Nevertheless, I . . . in elementary school, we had shop and I wasn't any great carpenter or anything. I had difficulty in getting something straight, or planing an edge properly or something.

T. Can you tell me more about what kind of a person that is?

P. The person doesn't look very smart, he looks sleepy, I guess. I . . . think that what comment I could make about both of these people—they both seem sexless. The only thing that differentiates these two—this one's (*Figure 2*) smaller than that one (*Figure 1*), which would possibly make me think of this as a woman (*Figure 2*), and this one's a man—has broad shoulders. That's supposed to be curly hair on that one (*points to Figure 2*) but it doesn't look like it. Looks like she needs a shave. But certainly you couldn't tell that it was a woman, looks more like a little boy. I think that if I could draw, I would have possibly done something else. In drawing a woman, I think if I could draw, I would have attempted to draw a profile which would make it more obvious that what I was drawing was a woman—than something like this.

T. It is hard to tell which is a man and which is a woman because you have a defect in drawing.

P. That would seem like an awfully stupid reason. It . . . I could say that sometimes I find it very difficult to tell whether I was a man or a woman, and that difficulty remains until today. When I find that in certain situations, the role I should be playing I could play more accurately as a woman than as a man. One of those situations occurred to me the other day—I remember wanting to remember to mention it. I guess it slipped my mind.

T. Both of these characters look to you to be sexless.

P. Yeh . . . Sometimes I've thought that if people were actually sexless, that would possibly be the easiest way to (*record unclear*) but on the other hand I sometimes thought that if I could act the same way toward people of both sexes, that would make things much easier . . . As I've said it has always been very hard for me to understand why it wasn't possible or permissible to have feelings of affection toward a male as it was to have those same feelings toward a female but . . . although it is (*word unclear*) I certainly would say it was possible to have those feelings but it isn't possible to express them the same way toward both people.

T. There is a complete denial of sexuality in these drawings.

P. Ummm—the only way in the drawings I attempted to indicate the figure's sex was the curly hair and I was going to try to give her better-formed legs . . . but I couldn't do it. I still would possibly be inclined to say that it might be a defect in my drawing ability. I said if I were able to draw better, I might have drawn something else.

T. You feel that if you could draw better it might have been possible to, if you had more skill or talent in drawing, it might have been possible to sketch in more definitely indications of sex.

P. . . . Yeh—I think that, as I said a second ago, I would have drawn the female in profile, if I was able to draw . . . but . . .

T. Had you been able to draw a profile it would have been more definitely a woman.

P. I think it would have looked more like a woman. . . . . I think that in my mind I can see what I would like to draw. I can see the differences in different pictures but when it comes to putting it on paper, it's not possible.

T. You can't act out your feelings.

P. . . . Well, again I've always thought of drawing as a skill or something that would have about as much connection with my feelings as the ability to sew or to knit or to do something of that sort.

T. I think that perhaps in looking at these drawings some of your comments have perhaps gone to the core of the problem here. Possibly the repressing of sexuality completely on the one hand versus acting the same toward men and women and perhaps even more, the most basic of all, is not knowing that you yourself are a man or a woman.

P. . . . That . . . second problem which you mentioned has been one which I guess I've spoken about in the past, and one which has troubled me particularly on many, many occasions. I think that . . .

T. It is not one that we've worked on much this year. We've managed to skirt around it pretty effectively. Pressing minor issues have prevented us from getting at that perhaps.

P. I think that one of the reasons that the Army is . . . could possibly be appealing is that it would show quite clearly whether I was a man or a woman.

Discussion

Perhaps the most immediately arresting characteristic of these drawings is that, in the words of the patient, "they both seem sexless." The patient's subsequent admissions offer strong evidence for the diagnostic meaning of the drawing of figures in which there is essentially no sexual differentiation.

However, for purposes of the present paper, the writer wishes to stress the effect upon the therapeutic process of employing this simple projective technique. After an essentially unproductive series of interviews, the therapeutic activity was once more directed toward a very fundamental problem in the patient; in fact, it might almost be said simply that therapeutic activity was reinstituted. In his own consideration of his drawings, he had brought clearly into the focus of therapy his basic confusion of sex identity. Although this confusion had been alluded to previously, he had now raised this problem in a different context with new and unavoidable implications. He had now made a start toward grasping the widespread ramifications that this confusion was producing in his interpersonal relations. Thus, there had been established, and by the patient himself, a basis with a concrete point of reference for the subsequent interviews.

In this one instance, the use of human figure drawings in the course of psychotherapy was highly, if not dramatically, successful. The writer does not purport to imply that such a result is an invariant. However, the technique has been of value in other cases. With another student, to cite briefly a different kind of usefulness, the drawings produced were extremely simple stick figures. It was suggested by the therapist that these drawings expressed compliance marked by minimal cooperation. This led directly to the patient's expression of annoyance at being requested to perform such a task, to a brief discussion of the similarity of the therapist's manner to that of the student's father, and, finally, to a profitable examination of the patient-therapist relationship.

Conclusion

Generalization from the single case is more than hazardous. However, the writer feels that there is ample evidence here to suggest, at the very least, consideration of using this projective technique within the therapeutic framework. This suggestion has its historical antecedents in the literature referred to in the introductory section of this paper. Although one may surmise that such a therapeutic use of drawings will be more limited in its effectiveness than the use of the TAT, it offers to be adequate to some of the problems in the course of therapy. Furthermore, the present technique has its own advantage. Since it is less time consuming in administration and since it obviates the necessity of preparing a transcription for the client, its use is more flexible within the interview.

References

1. Bettelheim, B. Self-interpretation of fantasy. *Amer. J. Orthopsychiat.*, 1947, 17, 80-100.
2. Deabler, H. L. The psychotherapeutic use of the Thematic Apperception Test. *J. clin. Psychol.*, 1947, 3, 246-252.
3. Gough, H. L. The frame of reference of the Thematic Apperception Test. *J. clin. Psychol.*, 1948, 4, 90-92.
4. Hartwell, S. W.; Hutt, M.; Andrew, G.; and Walton, R. E. The Michigan Picture Test: diagnostic and therapeutic possibilities of a new projective test in child guidance. *Amer. J. Orthopsychiat.*, 1951, 21, 124-137.
5. Jaques, E. The clinical use of the Thematic Apperception Test with soldiers. *J. abnorm. soc. Psychol.*, 1945, 40, 363-375.
6. Martinez-Arango, C. and Lasaga y Traviesо, J. I. Psychotherapy based on the Thematic Apperception Test. *Quart. Rev. Psychiat. Neurol.*, 1947, 2, 271-287.
7. Meyer, M. M. The direct use of projective techniques in psychotherapy. *J. proj. Tech.*, 1951, 15, 263-268.
8. Murray, H. A. Uses of the Thematic Apperception Test. *Amer. J. Psychiat.*, 1951, 107, 577-581.
9. Pepinsky, H. B. Application of informal projective methods in the counseling interview. *Educ. psychol. Measmt.*, 1947, 7, 135-140.
10. Symonds, P. M. Current trends and developments in the field of projective techniques. *Personality*, Symposium No. 2, 1-12. New York: Grune and Stratton, 1950.

SELF-PERCEPTIONS IN STUTTERING

SEYMOUR L. ZELEN,* JOSEPH G. SHEEHAN AND JAMES F. T. BUGENTAL

University of California, Los Angeles

Problem

A fundamental goal in the treatment of stuttering is the alteration of the stutterer's attitudes and perception of himself. Bryngelson [3], Johnson [5, 6], and Van Riper [13, 14] especially have stressed this point, finding the stutterer's attitude responsible for perpetuating the behavior. It is with an awareness of the importance of self-attitudes of the stutterer that the present study was undertaken.

Stutterers are customarily referred to and treated as handicapped persons. How much handicaps in general and stuttering specifically influence the self-concept of the individual, and thus behavior, is of major clinical import. A previous study by Sheehan and Zelen [7] on the level of aspiration of stutterers indicated that this self-concept has widespread effects. Therefore, it is most important to have some objective criteria of what the stutterer's self-percepts are, as well as some objective standards of comparison between the self-percepts of the stutterer and the normal speaker. Fiedler and Wepman [4] recently explored the self-concepts of ten stutterers employing the Q technique, finding no characteristic differences from non-stutterers.

Method

Using a more phenomenological and direct approach, the W-A-Y technique [2], the present study examined the self-perceptions of thirty stutterers. This group was then compared with a control group of 160 (all normal speakers) used in previous studies [1, 2].

In the W-A-Y technique the subject is told:

> "I don't know you very well, and so I wonder if you would do something that will help me to know you better. Take this piece of paper and write 3 answers to the question I will ask you. Remember, write 3 answers to the one question. You may write anything you wish—words, phrases, sentences—so long as you satisfy yourself that you have answered the question three times. The question is 'Who Are You?' "

The test situation was an early meeting of two stutterers' therapy groups. The question 'Who are you?' was asked of each group, and then the authors circulated among the subjects to assist with individual difficulties in responding. This aid simply consisted of restating and clarifying the directions. It should be noted that the authors were aware that this was a "classroom" type of situation, and that the structuring of segregation into this particular group might tend to create a set to have the self-perception "stutterer". Realizing this, they treated the response "stutterer" as merely another form of the group membership category.

This leads to a statement of categories or the dimensions of analysis of the responses. There were eleven in all:

Name
Familial relationship
Social positions of prestige
Occupation
Age
Sex
Group membership, including stutterer
Unit: pointing out uniqueness of the respondent
(e.g., myself, me, an individual)
Positive affect
Negative affect
Other—responses which do not fit into the other categories
(e.g., sports fan, athletic, 5 ft. 9 inches tall)

---

*Now at the Child Welfare Research Station, State University of Iowa.

Results

When the stutterers as a group were compared with the non-stuttering group, an over-all *F* ratio was obtained and found to be significant at the 5% level. Individually the only difference significant at the 1% level of confidence was on the "age" dimension, where the "controls" tended to respond with their age more than the stutterers.

The categories of self-percepts which differed at the 5% level of confidence were (a) *Positive affect*, with the stutterers showing more positive affect percepts than the normal core group; (b) *Unit*, with the normal group giving a significantly higher amount of unit percepts than the stutterers; (c) *Sex*, with the core group having more self-percepts in the sex-identification dimension; and (d) *Group membership*. The stuttering group responded with significantly more self-percepts involving group membership. However, when the percept "stutterer" is withheld from this category, the difference is not significant.

The comparison of the male and the female stutterers yielded no significant differences, but a comparison of the stutterers' groups divided on the basis of successful or unsuccessful termination of the course of treatment indicated only one significant difference: *Social status*, in which at the 5% level the more successful individuals tend to have a significantly greater number of self-percepts.

The *Negative affect* dimension yielded a difference between the successful and unsuccessful stutterers of better than 10% level. While this difference cannot be called significant, still it does serve to call attention to the possibility that the individuals who were able to cope successfully with their stuttering may tend to have less negative feelings about themselves than those stutterers who failed to meet adequately the challenge of this therapy.

The results of the present study are not to be thought of as contradictory to the findings of Fiedler and Wepman, but merely different. Although the differences reported here would not have appeared had we used 10 cases instead of 30, it should be pointed out that the dissimilarities in procedure and method of measurement are so great that the results cannot be directly compared.

Discussion

On 5 dimensions—age, positive affect, sex, unit, and group membership—our stutterers differed significantly from normal speakers in their self-percepts. Surprisingly, these stutterers had more positive feelings about themselves than the core group of normal speakers. This finding can be interpreted in either of two ways: (1) that stutterers tend to overcompensate for their feelings of inferiority and so protect themselves with a halo of positive feeling; (2) more fundamentally perhaps, that these particular stutterers, having decided to help themselves through psychotherapy, had resolved a major conflict in their phenomenal field and thus had more positive feelings about themselves. *Unit*, *age*, and *sex* self-percepts were lower in stutterers, and this may be interpreted as being a function of the stutterer's lessened concern with a census-type perceptual orientation and having instead a heightened orientation to his problems. Thus, stutterers have an increased responsiveness on the positive-affect dimension and the group-membership dimension, which included the percept "stutterer".

Rogers[11, 12], Raimy[9, 10], Snyder[8] and others have demonstrated that positive feelings increase and negative feelings decrease with successful psychotherapy, a finding with which the present results are in agreement. Since the successful stutterers tended to have fewer original negative-affect self-percepts, it is suggested that the W-A-Y technique might become useful as part of a battery for predicting response to therapy.

The significantly greater frequency of *Social Status* relationships in the self-percepts of successful patients provides an interesting suggestion of what might be one of the bases for favorable response to group treatment. Those patients who have strong needs for social status seem to have a greater capacity for group identification and thus are able to utilize this by involving themselves more deeply in the group.

Summary and Conclusions

The reported self-percepts of 30 stutterers were studied by means of the W-A-Y technique. Comparison of the stutterers with 160 non-stutterers yielded significant differences on these dimensions: age, positive affect, unit, sex, and group identification. Use of the last-named category was significantly greater for the stutterers because it included this self-percept—stutterer.

The use of the W-A-Y technique for predicting the stutterer's response to treatment is suggested by the observed tendency of the eventually successful patients to start therapy with fewer negative self-percepts. Possible implications of the results for group therapy with stutterers are discussed.

References

1. Bugental, James F. T. and Zelen, Seymour L. "Who are you?" A preliminary report on a method for studying the phenomenal self. *Amer. Psychologist*, 1949, *4*, 387. (Abstract)
2. Bugental, James F. T. and Zelen, Seymour L. Investigations into the 'self-concept'. I. The W-A-Y technique. *J. Personality*, 1950, *18*, 483-498.
3. Bryngelson, B., Chapman, M. E. and Hansen, O. K. *Know yourself: a workbook for those who stutter.* Minneapolis: Burgess Publishing Co., 1950.
4. Fiedler, Fred E. and Wepman, Joseph M. An exploratory investigation of the self-concept of stutterers. *J. Speech Hearing disorders*, 1951, *16*, 110-114.
5. Johnson, W. *People in quandaries: the semantics of personal adjustment.* New York: Harpers, 1946, 532 pp.
6. Johnson, W., Brown, S. F., Edney, C. W., Curtis, J. F. and Keaster, J. *Speech handicapped school children.* New York: Harper, 1948, 464 pp.
7. Sheehan, Joseph G., and Zelen, Seymour L. A level of aspiration study of stutterers. *Amer. Psychologist*, 1951, *6*, 500. (Abstract)
8. Snyder, W. U. (ed.). *Casebook of non-directive counseling.* Boston: Houghton Mifflin, 1947.
9. Raimy, Victor C. The self-concept as a factor in counseling and personality organization. Ohio State Univ. Abs. doct. dissert., 1944, No. 44, 99-105.
10. Raimy, Victor C. Self-reference in counseling interviews. *J. consult. Psychol.*, 1948, *12*, 153-163.
11. Rogers, Carl R. *Client-centered therapy.* Boston: Houghton Mifflin, 1951.
12. Rogers, Carl R. *Counseling and psychotherapy.* Boston: Houghton Mifflin, 1942.
13. Van Riper, C. *Speech Correction: principles and methods* (2nd ed.). New York: Prentice-Hall, 1947.
14. Van Riper, C. *Stuttering.* Chicago: National Society for Crippled Children and Adults, Inc., 11 S. LaSalle Street, 1948, 60 pp.

THE DIAGNOSTIC VALUE OF THE ROSENZWEIG P-F STUDY

ROBERT L. BROWN AND OLIVER L. LACEY

Furman University *University of Alabama*

Problem

The Rosenzweig Picture-Frustration Study, described elsewhere[1, 4, 5, 6], has gained some degree of popularity in the study of frustration at the human level. This projective technique was designed to assess reactions to frustration and is based on the author's observations and theory of frustration[2, 3]. According to Rosenzweig: "The mental disorder best illustrating each type of reaction would . . . be as follows: for the extrapunitive, paranoid complaints; for the intropunitive, compulsive and obsessional ones; and for the impunitive, hysterical ones".[3, p. 588] He has also given sample cases[4] of the patterns of responses found among normals and psychotics. Clearly, it is assumed that the technique has diagnostic value—that it differentiates in some way between those who are mentally and emotionally disturbed and those who are still able to function efficiently in society.

Does the Rosenzweig Picture-Frustration Study differentiate between various clinically labeled groups and, if so, does it differentiate to the extent that it can be used as a device in determining the diagnosis of unlabeled individuals?

Method

The P-F Study was administered to 108 subjects who were divided into three equal groups; a group of normal subjects, a group of chronic alcoholics, and a group of paranoid schizophrenics; with an equal number of males and females in each group. The normal groups were made up of university students and hospital personnel. The female alcoholic group was comprised of new members in Alcoholics Anonymous, while the male alcoholics were hospital patients diagnosed as "chronic alcoholic without psychosis". The subjects in the paranoid schizophrenic groups were selected by hospital personnel, on the basis of "clear diagnosis", recency of diagnosis, age, and education. All groups ranged in age from twenty to sixty years and in education from seven to seventeen years. The responses were scored according to the determinants set forth in the manual[6], and the numerical scores (percentages) were subjected to statistical analysis.

Results

There are three possible ways in which this test may be found to differentiate between the groups. (1) By the direction of aggression expressed; that is, whether it is extrapunitive, intropunitive, or impunitive. (2) By the type of response given; that is, whether it is ego-defensive, or is dominated by the obstacle, or shows a need to find a solution to the frustrating problem. (3) By the Group Conformity Rating which is obtained by comparing each record with a pattern of responses which has been found to occur in a sufficiently high number of records to permit its adoption as a norm[4, 6].

It will be recalled that the basic question raised is twofold: 1) does the P-F Study differentiate between variously labeled clinical groups, and, 2) does it differentiate between these groups to the extent that it can be used as a device in determining the diagnosis of unlabeled individuals?

The first part of the question will be considered by examining the results of this investigation (reported in Table 2) with special reference to the significance of the F

Table 1. Mean Scores on P-F Study for Six Groups of Eighteen Subjects Each and Standard Deviations of the Individual Scores

Scores	Groups					
	Normal		Alcoholic		Paranoid	
Direction of Aggression	M	F	M	F	M	F
Extrapunitive	35.3 (11.1)	36.5 (15.6)	36.4 (12.6)	49.4 (18.0)	24.6 (9.8)	39.3 (16.5)
Intropunitive	30.3 (6.4)	29.7 (10.3)	27.7 (5.5)	26.5 (11.3)	31.1 (10.0)	27.8 (9.8)
Impunitive	34.4 (7.4)	33.8 (10.4)	35.9 (10.1)	24.1 (8.0)	44.3 (9.3)	32.9 (10.2)
Type of Response						
Obstacle-Dominance	16.4 (6.7)	21.3 (9.3)	15.2 (4.9)	17.5 (7.5)	16.9 (5.9)	24.6 (8.6)
Ego-Defensive	51.4 (6.6)	51.0 (11.5)	54.1 (9.5)	59.1 (9.3)	53.8 (11.9)	53.4 (8.7)
Need-Persistence	32.2 (5.7)	27.6 (10.3)	30.8 (7.8)	23.4 (9.4)	29.3 (12.7)	22.0 (9.8)
Group Conformity Rating	66.5 (12.4)	69.3 (10.3)	63.7 (13.2)	62.5 (13.0)	60.9 (12.3)	60.2 (18.1)

TABLE 2. RESULTS OF THE F TEST ON SCORES DERIVED FROM THE P-F STUDY ON SIX GROUPS OF EIGHTEEN SUBJECTS EACH

	D.F.	E	I	M	O-D	E-D	N-P	GCR
Total	107							
Between Group	5	5.59**	.88	8.62**	4.26**	1.56	3.27**	1.19
Within Group	102							
Between Sex	1	12.42**	1.11	19.60**	12.36**	.57	12.13**	.006
Between Classification	2	5.42**	1.15	7.64**	3.30*	2.69	1.87	2.59
Discrepance	2	2.36	.25	4.10*	1.20	.93	.26	.37

*Significant at the 5% level
**Significant at or beyond the 1% level

scores between classifications, for the direction of aggression, type of response, and Group Conformity Rating. Significant differences between classifications in the percentages of extrapunitive and impunitive responses were found, both being significant beyond the one percent level. The only other significant class difference was in the percentage of obstacle dominance responses, which was significant at the five per cent level. In each case, the difference between sexes is of greater significance than the difference between classification. An examination of the mean scores (Table 1) for the groups shows that the source of the significant difference for the extrapunitive and impunitive responses lies in the deviations of the alcoholic female and paranoid male groups, with all other scores clustering closely around the normal. As to the significance of obstacle dominance responses (at 5% level), the mean scores show that the normal females and the paranoid females are slightly above the average for the groups, but are not more than one standard deviation of the raw scores away from the average. The Group Conformity Rating failed completely to differentiate between the groups in this study.

A quick survey of Table 1 immediately answers the second part of the basic question—does the P-F Study differentiate between various clinically labeled groups of persons to the extent that it can be used as a device for determining the diagnosis of unlabeled persons. The difference between the mean of any diagnostic group and the corresponding mean for any normal group is no greater than one standard deviation of the individual scores.

DISCUSSION

The results of this investigation show that the test has some discriminating capacity. However, the three significant differences which were found, do not appear impressive since there were no clear-cut differences between any one group and any other group. In view of this fact, we must conclude that these trends of differences are not discriminating enough to justify the use of the test for individual diagnosis.

Furthermore, it seems evident that the theoretical background of the test is not supported. As indicated earlier, Rosenzweig has suggested that paranoid complaints should be characterized by high extrapunitive scores, that compulsive and obsessional ones should be high in intropunitive scores, and that hysterical complaints should rank high in impunitive scores. In this investigation there is no group of hysterical disorders nor of compulsive disorders, unless the chronic alcoholic group may be considered as falling into the latter classification. In any case, the chronic alcoholic group did not score significantly higher than the normal or control group in either intropunitive or impunitive responses. The paranoid group, according to the theoretical prediction, should have scored higher on extrapunitive response

than either of the other groups. Actually, however, they gave a lower percentage of extrapunitive responses than did the normal and alcoholic groups. The paranoid males scored considerably lower than the normal males in extrapunitive responses and the paranoid females, although averaging a little higher, did not differ significantly from the normal females.

Summary and Conclusions

The P-F Study was administered to 108 subjects who fell into three equal groups; normals, chronic alcoholics, and paranoid schizophrenics; with an equal number of males and females in each group.

The results permit the following conclusions:

1. The test has some discriminatory capacity. However, the over-lapping of the group scoring is very great.

2. The differences between groups are not large enough to justify the use of the test in individual diagnosis. The difference between the mean for any diagnostic group and the corresponding mean for any normal group is no greater than one standard deviation of the individual scores.

3. The findings fail to confirm the theoretical validity of the test.

4. The P-F Study should not be used routinely in clinical diagnosis until further exploratory and experimental work yields an adequate validation.

References

1. Clarke, Helen Jane, Rosenzweig, S. and Fleming, Edith. The reliability of the scoring of the Rosenzweig Picture-Frustration Study. *J. clin. Psychol.*, 1947, 3, 364-370.
2. Hunt, J. McV. (Ed.) *Personality and behavior disorders. Vol. 1.* New York: The Ronald Press, 1944.
3. Murray, H. A. (Ed.) *Explorations in personality.* New York: Oxford University Press, 1938.
4. Rosenzweig, S. The Picture-Association Method and its application in a study of reactions to frustration. *J. Pers.*, 1945, 14, 3-23.
5. Rosenzweig, S., Clarke, Helen Jane, Garfield, Marjorie S. and Lehndorff, Annemarie, Scoring samples for the Rosenzweig Picture-Frustration Study. *J. Psychol.*, 1946, 21, 45-72.
6. Rosenzweig, S., Fleming, Edith and Clarke, Helen Jane. Revised scoring manual for the Rosenzweig Picture-Frustration Study. *J. Psychol.*, 1947, 24, 165-208.

CASE REPORT ON A MENTAL HYGIENE CLINIC PATIENT TESTED BEFORE AND AFTER FRONTAL LOBE INJURY[1]

LAWRENCE S. ROGERS[2] AND JAMES W. TAYLOR

Veterans Administration Regional Office
Denver, Colorado

Background Data

It is most unusual that test results are available on a non-psychotic patient, both prior and subsequent to serious brain injury. A 22-year old patient presented himself for treatment at the Veterans Administration Mental Hygiene Clinic in 1948 complaining of numbness and paralysis of his left arm. Previously a physical examination revealed no organic disease of the nervous system, and the E. E. G. was within normal range. This negative finding was confirmed at the time of his acceptance for treatment at the clinic. At that time, he was eligible for compensation for psychoneurosis, anxiety type, but had waived compensation in order to join the

[1]This paper was read at the annual meeting of the Colorado-Wyoming Academy of Science at Colorado Springs, Colorado, May 1, 1953.

[2]Our thanks are due to Dr. Johann R. Marx, Assistant Chief, Denver Veterans Administration Mental Hygiene Clinic for reviewing the medical material on this patient; and to Drs. Howard Siple and Lewis Bernstein, of the Veterans Administration Hospital, Denver, for making some of the test results available.

naval reserve. He was described by his therapist as having marked mood swings, being highly suggestible, immature and impulsive. His behavior was described as being loud and boisterous, obviously an attention getting device. After five treatment interviews, mainly supportive in nature, his arm returned to normal functioning. His therapist attributed the change to his extreme suggestibility. However, in the course of treatment, he was able to discuss his hostility towards his father, towards his employer and about his reaction to military experiences. The closing diagnosis was conversion reaction. While in treatment he was given the Rorschach and Wechsler-Bellevue.

Subsequently, the patient attended college for one and a half years. His adjustment was adequate, although he complained of nervousness and had some difficulty with school work. On a routine compensation examination, his behavior was described in the same terms as when he was a clinic patient. However, he had no complaints of paralysis.

The patient was called back to active duty by the Navy and had some service in the Korean area. In April, 1951, the veteran suffered a frontal skull fracture and frontal lobe injury in an automobile accident. He was completely unconscious for several days and consciousness was regained gradually. Eight days after his injury, he was able to respond only to a slight degree. He had a complete right hemiparesis and pronounced aphasia.

His recovery continued to be gradual. Seven weeks after his accident, the aphasia was gradually subsiding but it was still difficult for the patient to choose the correct words. At that time, there was a marked right spastic hemiplegia, only the shoulder and arm muscles could be moved. The muscles of the fingers, leg and foot were nonfunctioning. There was also a marked lessened sensitivity to pain.

The next report is dated 7 months subsequent to the injury. It was noted that movement was restored to the right extremities but that there was marked incoordination with moderate atrophy of the muscles. Reflexes were hyperactive on the right side. The patient was described as mildly aphasic, mildly euphoric with a marked loss of recent memory. The diagnosis of encephalopathy was confirmed. The E. E. G. was abnormal, showing an irregular slowing which would be consistent with a diagnosis of convulsive disorder. The patient was put on anticonvulsive medication as a prophylactic measure since there were no reported seizures. During the period of hospitalization, the patient was given the Wechsler-Bellevue Scale, and the Wechsler Memory Scale, among others.

The patient subsequently returned to treatment at the Veterans Administration Mental Hygiene Clinic for psychotherapy and anticonvulsive medication. He complained of headaches, and inability to move his right arm above the level of his shoulder. It should be remembered that he had complained of an inability to move his *left* arm at the time of previous treatment. He also complained of a memory defect and difficulty in speech. He was described as being very immature, exhibiting flighty behavior, being quite boisterous and emotionally very labile. Essentially, his behavior appeared to be an exaggeration of the type described prior to his injury. The patient has been seen in treatment for 12 months, during which time he received both psychotherapy and anticonvulsive medication. His mother was also seen three times and she is described as a very dominating and overprotective woman. The patient also had been referred to the Denver University Speech and Hearing Clinic and received considerable help from this source for his aphasic condition.

This patient has shown considerable improvement; his flighty behavior, silly talk, and boisterousness have all diminished considerably. His complaints of memory loss, aphasia and inability to move his arm have disappeared. He obtained a position and has shown a sense of responsibility and a capacity for work which did not exist prior to his injury. He is still being carried on in therapy, mainly for the anticonvulsive medication and supportive psychotherapeutic interviews. The patient was tested during this period of treatment in March, 1952, 11 months after his injury.

Test Results

Wechsler-Bellevue Intelligence Scale. The weighted scores on each of the Wechsler-Bellevue subtests, as well as the Verbal IQ, Performance IQ, Total IQ and Deterioration Quotent (DQ) are presented in Table 1.

Table 1. Wechsler-Bellevue Weighted Scores and IQ's Before and After Head Injury

W-B Subtests and IQ Scores	Wechsler's "Organic" Signs	33 Months Prior	7 Months After	11 Months After
I	+	11	10	11
C	+	9	9	8
D	- -	14	7	11
A	-	10	9	13
S	-	12	8	11
V	++	9	12	10
PA	0	9	12	12
PC	0	12	12	13
BD	- - to 0	13	10	15
OA	0 to - -	13	12	14
DS	- -	13	10	11
VSIQ	V>P	109	99	107
PSIQ		113	110	122
FSIQ		111	105	116
Hold		45	46	48
Don't Hold		50	36	50
D. Q.		-	22%	-

Quantitative analysis of each of these tests, taken individually, would not indicate organic brain damage. The only testing which gave a DQ which might be suspicious was 7 months after his accident. But even here, the DQ is only 22%; a finding not unusual in the clinic "non-organic" population [6]. At this time, the patient's profile revealed three of Wechsler's "Organic Signs", but neither Block Design nor Digit Symbol differed significantly from the performance mean score. Also, the Performance scale IQ is still much higher than the Verbal scale IQ, again contrary to Wechsler's expectations.

Eleven months after the injury, the only "organic" sign is the low Digit Symbol test score. The DQ is negative, and Block Design is the highest of all subtests. The Verbal IQ is still much lower than the Performance IQ.

The Reynell [5] index for organicity was computed for each test. The index was negative for the first and third tests, but positive (14 points) for the test at 7 months. The Hewson [1] ratios were also computed. These were remarkably consistent; for each of the tests the same two of the eight ratios were positive. Apparently the brain injury had no effect on these ratios.

However, while analyses of the individual tests were not revealing, a comparison of the tests is most rewarding. Comparing the test results at 7 months after injury with the test prior to injury, we find that 9 of the 11 subtest scores have changed in the direction expected by the Wechsler signs. There is a marked decrement in capacity for memory (D), abstraction (S), perception and synthesizing (BD), and new learning (DS). The improvement in Vocabulary and Picture Arrangement scores seems to be related, in part, at least, to the patient's attitude and co-operation, since on the first test he manifested behavior consistent with the personality picture described above. On the second testing, he was extremely cooperative. In terms of IQ, it can be noted that his performance IQ is approximately the same. On the other hand, there is a marked decrease in the verbal IQ which is consistent with the aphasia. In addition, the sum of the "Hold" tests has remained consistent, whereas the "Don't Hold" tests have dropped 14 points.

In contrast to the comparison of the first and second testings, the first and third testings show little consistency with "organic" expectations. Memory capacity (D)

and new learning (DS) still show some decrement, although less than on the second test. Arithmetic and Block Design, which were lower on the second than the first test, now surpass the scores prior to injury. A comparison of the second with the third test indicates that most of the improvement was found in the "Don't Hold" tests. It is likely that some of these increases are due, in part, to practice effect since there was only a 4 months interval between the last tests. On the other hand, the better performance may be due to a reduction of the effects of the damage and return to the pre-injury level. This interpretation of the change would be consistent with the clinical impressions at the time. We were aware of the possibility that there might be some practice effect on the last testing; however, the advantages of comparison on the same material outweighed the disadvantages.

Rorschach. The patient was given the Rorschach only at the first and third testing. The scoring and tabulation are presented in Table 2.

TABLE 2. RORSCHACH SCORING AND TABULATION

Card No.	33 Months Prior	11 Months After
I	W F + A P	W F + A P D F + Hd D F - A
II	W F + A	D F + Ad P
III	Rejects	D M + H P D F + Cg
IV	W C'F - A	Rejects
V	W FM + A P	W F + A P
VI	Rejects	Rejects
VII	Rejects	Rejects (D F + Hd)
VIII	Rejects	D FM + A P D F + A
IX	Rejects	D F + Hd P D F ± At
X	Rejects	Rejects (D F + A)
Scoring	R = 4 Rejects = 6 W = 4 F + = 2 FM + = 1 C'F- = 1 A% = 100 P = 2	R = 11 + 2 Rejects = 4 W = 2 D = 9 F + = 7 + 2 F ± = 1 F- = 1 M + = 1 FM + = 1 At = 1 Cg = 1 H + Hd = 3 A% = 55 P = 5

It will be noted that this veteran gave only four responses the first time and eleven the second time he took the Rorschach. Detailed perusal of the second record failed to reveal the usual signs of the effect of the accident; neither Piotrowski's[4] nor Hughes'[2] lists of "organic" signs gave any clue to the fact that this man had had a frontal lobe injury within the year.

At the last testing, the Rorschach gives evidence of the patient's attempt at controlling his behavior and channeling his behavior into more acceptable avenues. Clinically, this striving for control was also observed. There seemed to be a wavering between his boisterousness at this time and his rigid attempts to control himself.

Wechsler Memory Scale. The Wechsler Memory Scale was also given 7 months and 11 months after the accident. In contrast to the Wechsler-Bellevue results, there was little change during this period. The Memory Quotient went up from 90 to 94. The change was due almost entirely to the increased digit span which was also in evidence on the Wechsler-Bellevue.

Discussion and Summary

The literature in this field is not at all helpful, primarily because the subjects used are either the hopelessly psychotic, the very severe neurotics or patients with intractable pain. Usually the opportunities for adequate preoperative testing are obviously extremely limited. Struckett[7] quotes 13 studies in this field: 4 reported a loss in intelligence test scores, 4 reported no gross decrease, 2 reported an increase, while 3 observed no significant difference. Reasons for the lack of agreement can readily be found including: the selection of subjects, operative approaches, absence of adequate controls, etc.

Many studies, including the Columbia-Greystone[3] and the Struckett studies[7] indicate that the greatest loss occurs immediately after the operation or injury, and that in 6 months or so the loss is negligable. Our patient did not follow this pattern. Of course, we do not have tests for the period immediately following the injury. But at 7 months he still showed very pronounced losses. It was only at 11 months after injury that he had surpassed his previous level.

This paper emphasizes the fact that the effects of organic brain damage do not always show up in the expected manner. We cannot assume that all types of injury to the brain will result in similar test behavior. Nevertheless, the clinician often functions as if this were true, and as if the effects were permanent. Yet, there are many studies, in both the human and animal field which indicate that much brain tissue can be lost without having a marked prolonged effect. However, there apparently are more subtle cues which the psychologist uses which have not yet been quantified to a degree that can be used as "signs."

Another major difficulty is the fact that we usually do not know the person's pre-injury level, either in personality or intellectual functioning. That is, we cannot assume that every patient is normal to start with, and that the deviations will all be from the same level. In this particular case, the patient did change in the expected manner on the Wechsler-Bellevue but the only way we could have known about it was by having the original test. On the Wechsler-Memory Scale, while there is a slight increase from the second to the third test, the patient's starting point is unknown. While the Rorschach was consistent with the clinical impression of the personality, the test did not contribute to the diagnosis of organicity. However, other tests at the last testing along with behavioral cues did indicate the possibility of organic brain damage.

References

1. Hewson, Louise R. The Wechsler-Bellevue Scale and the Substitution Test as Aids in Neuropsychiatric Diagnosis. *J. nerv. ment. Dis.*, 1949, 109, 158-183, 246-266.
2. Hughes, R. M. Rorschach Signs for the Diagnosis of Organic Pathology. *Rorsch. Res. Exch. & J. proj. Tech.*, 1948, 3, 165-167.
3. Mettler, F. A. et al. *Selective Partial Ablation of the Frontal Cortex.* New York: Hoeber, 1949.
4. Piotrowski, Z. A. The Rorschach Ink-Blot Method in Organic Disturbances of the Central Nervous System. *J. nerv. ment. Dis.*, 1937, 86, 525-537.
5. Reynell, W. R. A Psychometric Method of Determining Intellectual Loss Following Head Injury. *J. ment. Sci.* 1944, 90, 710-719.
6. Rogers, L. S. A comparative evaluation of the Wechsler-Bellevue Mental Deterioration Index for various adult groups. *J. clin. Psychol.*, 1950, 6, 199-202.
7. Struckett, Pauline B. A. Effect of Prefrontal Lobotomy on Intellectual Functioning in Chronic Schizophrenia. *Arch. Neurol. Psychiat.*, 1953, 69, 293-304.

RORSCHACH PATTERN ANALYSIS IN SCHIZOPHRENIA: A CROSS-VALIDATION STUDY[1]

EARL S. TAULBEE and BOYD D. SISSON

Norfolk State Hospital, Norfolk, Nebraska *University of Nebraska*

PROBLEM

In a recent publication Thiesen [3] reported five patterns of Rorschach variables which he found to be associated with schizophrenic patients and nonoccurring or extremely rare in normal subjects. While these patterns significantly differentiated the two groups, it was stated that they must be regarded as tentative inasmuch as the groups on which the patterns were tested were the same groups from which they were derived, and therefore needed to be confirmed on new groups. The following study reports the results of a cross-validation of the applicability of the patterns using ing new patient and control groups.

TABLE 1. PATTERNS ASSOCIATED WITH SCHIZOPHRENIA

Pattern	Description	Thiesen		Chi Square	P	Present Study		P*
		Controls (N 157)	Patients (N 60)			Controls (N 157)	Patients (N 62)	
A	An: High[1] Sex: High	1	12	25.5	.001	0	5	<.002
B	F+%:Low Z Score: Low	2	11	19.4	.001	2	3	.14
C	FC+: Low M: Low A%: Low	0	6	12.2	.001	2	5	.02
D	F+%:Low P: Low A%: Low	1	9	16.9	.001	3	5	.04
E	DW: High FC+: Low	1	7	12.0	.001	0	0	—

*Because of the low cell frequencies involved, Fisher's "exact test" was used for determining the probabilities.

1. An: High (3 and above—excluding sex responses)
Sex: High (1 and above)
F+%: Low (0-68.9)
Z score: Low (0-7.5)
FC+: Low (0)
M: Low (0)
A%: Low (0-33.9)
P: Low (0-4)
DW: High (1 and above)

[1]From the Veterans Administration Hospital, Knoxville, Iowa.

Subjects

Subjects used in this study were 62 hospitalized adult male schizophrenic patients, so diagnosed by the psychiatric staff of Knoxville, Iowa, Veterans Administration Hospital. The control group consisted of 157 "normal" men and women from a small midwestern community,[2] which is located in the same general geographical locale from which the patients came. The age range for the patient group was 20-41, mean age 28, SD 5.16; for the control group the range was 20-45, mean age 33.8, SD 6.89.

Procedure

The Rorschach test was individually administered to all subjects in both groups and scored according to the revised Beck system (1). Each record was examined for the presence of one or more of the five patterns described in Table 1. Derivation and further description of these patterns are contained in the previous paper (3).

Results

The assumption of homogeneity with respect to total pattern occurrence for the two groups was tested by chi square, employing Yates' correction for continuity, and rejected at the .001 level of confidence.

With respect to the individual patterns it is shown in Table 1 that the discriminating efficiency of Patterns A, C, and D is confirmed, whereas Patterns B and E fail to meet the acceptable statistical criterion.

In the present study only one multiple pattern, a C-D combination in the schizophrenic group, was obtained; in contrast, the earlier findings reveal that eight of the patient group gave two or more patterns.

The patient group consisted entirely of male subjects, whereas the control group was composed of 103 females and 54 males. A chi square of .079 between sex and pattern occurrence, however, allows inference that pattern discrimination for one sex is not significantly larger than for the other.

Discussion

The authors are in agreement with the interpretation offered by Thiesen that Pattern A is suggestive of "somatic preoccupation and release from culturally determined inhibition," which are not unlikely characteristics of a schizophrenic disturbance. The possibility that this pattern might be more common in selected populations of specialized interests or sophistication in the absence of psychopathology, as mentioned in the earlier study, remains unanswered, but it did not occur in any of the present control group. Likewise, it is agreed that Pattern D suggests a disturbance in affective contact with the environment. The confirmation of Pattern C offers no additional explanation of the interrelationship among the factors comprising this pattern.

Failure to confirm Patterns B and E may be attributed in part to the lack of homogeneity with respect to pattern occurrence in the two schizophrenic groups as evidenced by a chi square of 4.82 ($P = .03$). A suggested explanation for this difference is that the present patients were new admissions whose psychoses were in early stages of development or recurrence, and one half of whom were diagnosed schizophrenic reaction, paranoid type. For this group the mean Wechsler-Bellevue IQ was 100.65 (obtained within two days of Rorschach administration), and the mean F+% was 72. Also, compared with Thiesen's "normal" group, the present patient group obtained a significantly higher Z score and there were no significant differences between these two groups for A% and number of FC responses. While it is obvious that these factors are in opposition to some of those appearing in the patterns not validated, it is felt that their occurrence among schizophrenic patients where the

[2]The records were taken from a larger study done under the supervision of Dr. Marshall R. Jones and supported by the University of Nebraska Research Council.

illness is not advanced and among those of the paranoid type is not uncommon. There were no significant differences between the two control groups for the presence of patterns.

Summary

In a previous study Thiesen reported five Rorschach patterns associated with schizophrenia. The present study presents the results of a cross-validation based on a group of 62 adult schizophrenics and a group of 157 "normal" controls. Three of the patterns were found to be discriminating. A probable explanation is offered for the failure of the remaining patterns to be validated.

The authors believe that these results lend support to Thiesen's statement that when the patterns occur they may have clinical usefulness in early detection of a schizophrenic process or as an aid to determining the gravity of an initial psychotic episode.

References

1. Beck, S. J. *Rorschach's test. Vol. I. Basic processes.* (2nd Ed.) New York: Grune & Stratton, 1949.
2. Beck, S. J., Rabin, A. I., Thiesen, J. W., Molish, H., and Thetford, W. N. The normal personality as projected in the Rorschach test. *J. Psychol.*, 1950, 30, 241-298.
3. Thiesen, J. W. A pattern analysis of structural characteristics of the Rorschach test in schizophrenia. *J. consult. Psychol.*, 1952, 16, 365-370.

A MECHANICAL SHORT-CUT FOR COMPUTING TETRACHORIC CORRELATIONS

ANDERS SWEETLAND AND RALPH MASON DREGER

Florida State University

Introduction

As research accumulates in clinical psychology, the complexity of clinical problems becomes more and more apparent. In view of this complexity, more psychologists are looking towards factor analysis, as it seems to be a tool of sufficient stature to encompass the dimensions of clinical problems.

At first glance factor analysis is discouraging: the computational procedures are complex and time consuming; but even more forbidding is the number of correlations involved. An innocent-appearing study involving twenty elements necessitates the computation of 190 correlations. Even the thought of this can be traumatic, particularly since most psychologists do not have IBM facilities. However, if the problem of mass production of correlations could be resolved, factor analysis might become accessible to many psychologists who have previously been discouraged from using it.

In attempting to solve this problem we set up three desiderata: (1) That the method be inexpensive; (2) that it be reasonably economical of time; and, (3) that it be simple enough to be handled by a student assistant (who is frequently a college sophomore without mathematical sophistication.)

To reduce computational time, the first consideration is generally that of using tetrachoric correlations. However, the reduction of the groupings from the number in a scatter diagram results in a loss of accuracy: as the number of intervals is reduced the standard error increases. This objection can be surmounted by increasing the size of the sample, thus decreasing the error. But to increase the sample size involves additional clerical work, so the psychologist is back where he started. What is needed is a means of increasing sample size without an appreciable increase in clerical work.

Procedure

After formulating and rejecting several possibilities we decided to use punched cards for sorting of items and to weigh rather than count cards. (Any balance sensitive to 0.1 gram could be used.) After the cards were sorted into four categories as required for the tetrachoric r, each stack of cards was weighed and the weights used in the computations. A further reduction in the size of the standard errors was obtained by using a sorting division close to the mean of the sample for each item or test.

The following is a description of the testing of the procedure. An attitude scale of fourteen items was used with each item subdivided into an 8-point scale. The scale was administered to 350 college students. A sample of 75 scales was tabulated. After means were computed for the fourteen items, a template was prepared which exposed only the items above the mean in each case. Each individual's scores were then encoded on a 3″ x 5″ card. These cards were made by clamping them between two boards and drilling 14 holes around the edges, spaced approximately 5/8 of an inch apart, ¼ inch from the edge. Using a drill press speed of 1000 r.p. m., and counter-boring with a slightly larger drill (#3 drill counterbored with a #1 drill), we could obtain holes with practically no burring. Clean holes are necessary for ease in sorting. One corner of each card was clipped for purposes of alignment.

In scoring, the student assistant placed the template on the scale sheet, and, with a scissors, clipped out the edge of the corresponding hole on the card of any checked item showing. These operations were rechecked by another student for accuracy.

A knitting needle, made from a coat hanger, was used to sort the cards. The needle was first run through the item 1 holes and lifted up. Items that were answered "yes" (i.e., above the mean of item 1) fell to the table and the "no's" were left hanging on the needle. With these two stacks separated, the same procedure was followed for item 2, each of the item 1 stacks being sub-divided into above mean and below mean of the item 2 groups. The four stacks of cards obtained (item 1, yes —2, yes; item 1, no —2, no; item 2, yes —item 1, no; item 2, no —1, yes) were weighed on the balance. The cards were restacked, and item 1 and item 3 sorted. Again the four stacks were weighed. A similar process was followed to obtain the 91 correlations.

Table 1. Distribution of 90 Correlations in Total Grams Obtained by Adding the Four Quadrants

Grams	Frequency
512	1*
511	0
510	0
509	0
508	2*
507	0
506	0
505	2
504	2
503	5
502	6
501	8
500	9
499	16
498	20
497	8
496	7
495	2
494	0
493	0
492	2*
491	0

*Might be reweighed for possible clerical errors

With these weights tabulated, the tetrad of cells (or quadrants) were added by rows and columns and the correlations computed using Hayes'[2] charts. The Chesire-Saffir-Thurstone[1] tables or the cosine pi method could also be employed. By use of the Hayes' charts and a slide rule, computations take approximately two minutes.

A check on the accuracy of this technique can be made by glancing at the table of total weights. Any aberrant measures are immediately apparent. An idea of the accuracy of sorting can be seen in Table 1 which shows the total weights obtained by adding the four quadrants. The error is less than one per cent in 94% of the cases.

In checking to see how well we had met the desiderata, a final "expense sheet" was compiled: (1) total expense was sixty cents for four hundred 3″ x 5″ cards, plus seventy-five cents per hour for the assistants; (2) the *total time* consumed (including preparation for correlations) was approximately forty minutes per correlation. (Since the experimental trial included necessary delays, we believe that the time could be halved by a more organized approach); (3) an average error resulted of less than one per cent.

It is possible that the time might be further reduced by measuring the stacks of cards with a micrometer. We did not follow this practice as the 3″ by 5″ cards tend to distort in handling. The result is a "spongy" stack which is not suitable for micrometer measurements. Commercial cards (McBee, Keysort) would be much less prone to this shortcoming.

Briefly, the procedure is:

1. Determine means of items from sample of total N.
2. Make template(s) cut out above the mean.
3. Place template(s) on each scale, cut edges of holes where checked items show through.
4. Stack all cards and divide into four stacks with needle.
5. Weigh cards in each stack and record weights in 2 x 2 table.
6. Repeat #4 and #5 for each item.
7. Compute tetrachoric r's by diagrams.

Summary

A short cut for computing tetrachoric correlations involving the use of card sorting and weighing is described. Total cost for a sample involving 350 cases involving 91 correlations was less than one dollar for materials. The time was less than 40 minutes per correlation and the reliability of scoring showed an error of less than one per cent.

Bibliography

1. Chesire, L., Saffir, H., and Thurstone, L. L. *Computing diagrams for the tetrachoric correlation coefficient.* Chicago: Univ. of Chicago Bookstore, 1933.
2. Hayes, S. P. Diagrams for computing tetrachoric correlation coefficients from percentage differences. *Psychometrica,* 1946, 11, 163-172.

THE LEARNING PLATEAU SEEN AS RESISTANCE IN INDIVIDUAL REMEDIAL TEACHING

MONROE G. GOTTSEGEN AND GLORIA B. GOTTSEGEN

Inter Agency Guidance Center, Yonkers, N. Y.

The Developmental School, New York, N. Y.

Introduction

It is our experience in work with children referred for remedial reading, that one of the most significant occurrences which can act to negate achievement is the learning plateau. This is most apparent in individual remedial sessions when the child possesses an emotional difficulty. However, even with the normal child, whose referral seems essentially due to school failure of a non-neurotic nature, the learning plateau occurs to interfere with achievement[3]. The generalized blocking of the plateau may reflect the displacement of unconscious resistances seen as negative attitudes toward general as well as specific learning so that failure to progress can be better understood in terms of the emotional concomitants which underlie the structure and relationships of the remedial hour[1].

It is generally accepted in individual remedial teaching that a necessity exists for distinguishing between structure and permissiveness[2]. The degree of structure within the learning situation is determined by the original diagnosis of the specific reading handicap, coupled with individual capacity for learning, and based upon the emotional strengths and weaknesses of the child. In working with the child, keeping specific goals in mind, it is necessary to plan carefully the selection and use of materials. On the basis of the educational diagnosis made, the remedial teacher must allot time for work on specific areas necessary for correcting apparent disabilities, i.e., a typical remedial reading hour may consist of oral reading (10 minutes), work with related word recognition techniques and games (15 minutes), comprehension exercises (10 minutes), and story reading (10 minutes).

The time allotted to such activity is adhered to with a rather consistent degree of regularity. Permissiveness is shown during the activity undertaken in terms of the teacher-pupil relationship. It is of note that whether or not the remedial reading is being given while the child is undergoing psychiatric treatment, the therapeutically-oriented teacher nevertheless remains a teacher, and does not attempt to deal directly with observed attitudes and underlying affect. In the kind of remedial hour described, it is not unusual to find, after an initial learning spurt, a tendency for the child to block, reach a learning plateau, and then to regress. This seems to occur in spite of the attempt to vary materials within the hour, introduce new activities of a related nature, and thereby continually stimulate interest.

When blocking initially appears, it begins to show itself as a fairly isolated function of one specific learning area. However, the tendency then is for the blocking and related inefficient educational functioning to spread to the extent that it encompasses a majority of reading areas and skills. The child at the same time may be coming late or showing other forms of resistance of a more generalized nature. Upon conferring with parent and/or therapist, there may appear no overt indication upon which to base an explanation of the learning block. The plateau reached appears as if out of nowhere and, if ignored, can retard learning for a considerable period of time. It is the duty of the remedial reading teacher to exert all effort in overcoming the learning block. In many cases, successful remedial tutoring must rest upon the basis of whether the impinging regressive phenomena are dealt with and overcome.

After the initial spurt which may be due to expectations of quick progress on the part of the child, a wish to please parents and friends, or an initial fascination with the techniques and procedures useful in the hour, the novelty and fantasy expectations of quick success begin to wear thin when confronted with the reality of a

learning situation which depends for its success upon productive achievement. The child is aware, no matter how permissive and accepting is his teacher, that he is expected to learn in the hour, and to apply effectively those skills and techniques mastered, to school and home situations. Whatever the attitudes originally interfering with productive learning in school, they are at this point partially displaced in the form of resistances to the remedial hour. Prior success within the reading hour, contrary to expectation, is not sufficiently supportive for overcoming the resulting plateau.

A hostile child who has found it difficult to be directly aggressive in his relationship with parents and who has not adjusted in peer relationships may transfer the latent negative attitudes toward authority (as seen in failure to apply sufficient energy for academic achievement) to the remedial teacher. If, in continuing the above example, negativity appears as an expression of hostility in school endeavour, it will also make a similar appearance in the tutoring hour. It is important to realize that the permissiveness of the remedial teacher will encourage the display of the defensive mechanisms and increase tendencies to displace and project. Again, a youngster who is known to be passive and withdrawn, deriving major satisfactions from a wish-fulfilling fantasy, will dawdle, procrastinate, lose his place, and function with extensive inefficiency under similar circumstances. If the remedial teacher anxiously reacts with a lessening of permissiveness and a more rigid adherence to structure, he will create further anxiety in the child, rather than dispel it, and as a result totally destroy the opportunity for a reorientation toward new learning. If he ignores the resulting phenomena inherent in the seemingly unexplained failure to learn, and in the behavioral expression of resistances, he will find himself frustrated in his teaching and of little help to the child. The remedial teacher cannot ignore the presence of a resistance simply because the pupil may be receiving psychotherapeutic treatment elsewhere. The child himself basically desires that his behavior be dealt with constructively, without having, as in the past, to suffer conflict stemming from a reactive anxiety on the part of the school and home.

Case Studies

Two cases follow in which resistances were encountered and dealt with successfully. The first is that of a pre-psychotic boy who was receiving play therapy as well as individual remedial reading. The second concerns a youngster who exhibited no observable abnormal processes, either in the nature of the affective controls, the quality of conceptualization, or the manner of relating to peers and adults.

Vince, aged 9, tested at the borderline level of intelligence on the Wechsler Intelligence Scale for Children. He was considered a culturally handicapped youngster whose ability to learn was complicated by a pre-psychotic emotional orientation. A neurotic defensive attitude was employed by Vince to deny his deep doubts concerning his failure to learn or compete with classmates. As a result of being dominated, yet rejected, by his older brother who used every means possible to destroy Vince's self-esteem, Vince possessed a sincere belief that he was entirely incompetent, stupid and defective. In defending himself against experiences which could produce anxiety, he reacted by refusing to attempt any productive goal-seeking activity, and in doing so, became hyperactive and facetious, clowning his way out of anxiety-arousing learning situations and relationships. Although overtly conforming, the behavior exhibited showed indirect evidences of intense hostility expressed in childish evasiveness and negativity. Vince's purposely obnoxious behavior inevitably aroused the hostility of his teachers.

As a result Vince was seen for play therapy on a once-a-week basis, and for remedial reading, two hours per week. Although the behavior described above did not appear during the first few hours of the remedial sessions, it nevertheless shortly occurred. Resistances took the form similar to the kind described above. Dealing with the resistances meant almost completely doing away with the minimum amount of structure involved in the initial reading sessions. Vince was read to, for approximately twenty minutes; simple games and other reading activities of a readiness nature were depended upon exclusively with the central purpose being to entertain and amuse under the guise of teaching reading. The purpose here was to allay anxiety by permitting the rationalization of achievement, to build relationship and stimulate identification, and to make unnecessary the employment of defensive behavior as seen in the child's approach and attitude toward new learning. When anxiety had decreased sufficiently, the defenses could be relaxed, and it was possible to introduce a more direct teaching approach.

Dick, aged 8, achieved a classification of low-average intelligence on the Stanford Binet L examination. He was, at the time of referral, functioning at middle to upper first grade level in his reading. His work habits were characterized by extensive sounding out of words, by exclusive use of a phonetic approach, and a deliberate attempt to solve rather than to read material presented to him. There was no indication that Dick had ever grasped the concept of reading in phrases. He was the perfect example of the word by word reader who was fascinated by the results of his analytical approach. Within a short period of time, on the basis of two hours per week, Dick gave evidence of being able to increase his word recognition abilities to the extent that he was shortly reading in meaningful phrases. He was able to take obvious pleasure in free reading assignments which stimulated his imagination.

Following this initial achievement, in which his reading efficiency showed a rapid and effective increase, he was observed to become extremely careless in his word recognition, demonstrated a tendency to lose his place with more than a fair degree of regularity, and to be overly concerned with illustrations and game materials for which he had never in the past shown a great liking. The difficulties first appearing in the oral reading were shortly transferred and extended to the word recognition exercises and games. When the remedial teacher became aware that the learning inefficiencies and growing negativity were playing an increasingly dominant role, attempts were made to utilize, extend, and to give a more prominent place to, the pleasurable parts of the reading hour. The remedial teacher spent a goodly portion of the time in reading stories to Dick. He increased and rotated the game material on the theory that the correlated oral reading activity permitted a greater opportunity for concretizing resistances and generating anxieties. Oral reading was cut to minimum and eliminated for quite a few sessions. After eight sessions the efforts of the teacher to counter resistance by changing the character of the authority-initiated routine, creating a further positive basis for relationship, and not extending the area of conflict by responding with anxiety, permitted a slow reintroduction of some of the resisted tasks, thereby gradually overcoming the learning plateau. The remedial teacher was, in effect, attempting to show Dick by deed resulting from empathic responsiveness, that the displaced resistances were untenable in terms of the reality relationship; that by being able to demonstrably modify the basis upon which the teacher-pupil relationship was formed, it was possible for Dick to acknowledge flexibility in attitudes toward authority relationships. In short Dick was given observable emotional support for altering his distorted attitudes toward his teacher, who was originally conceived of as an accepting, but nevertheless authority, figure.

Because a youngster's attitudes toward adults are derived from relationships in school and family, it is unavoidable that the child views future adult contact in terms of past contacts. Experience shows that permissive conditions encourage projection and displacement of feeling and attitude. It is therefore apparent why the child, anxious about his failure in school, and tending to show evidence of the denial of his own responsibility by developing, through rationalization, negative attitudes toward teacher and other authority figures who demand achievement, will, when permitted, overtly display related affect and attitude under conditions which permit their expression. In effect, learning plateaus evidenced in the remedial hour seem to be an expression of hostility toward authority adults, who play and who have played a major role in the life of the child.

Summary

Two remedial reading cases, one of an emotionally disturbed and the other of a normal boy, were cited in order to show how resistances appearing in the form of plateaus occur and how they may be dealt with by the educational psychologist within the remedial reading hour. It is concluded that the teacher himself has the responsibility to recognize the plateau when it occurs, understand it in terms of emotionally related attitudes governing approaches to new learning, and attempt to deal with it through manipulation of the relationship and learning environment.

References

1. Freud, Anna. *The Ego and The Mechanisms of Defense.* London: Hogarth Press, Ltd. 1947. 3rd impression.
2. Harris, Albert J. *How to Increase Reading Ability.* New York: Longmans, Green & Co., 1948.
3. Pressey, Sidney L. and Robinson, Francis P. *Psychology and The New Education.* New York: Harper & Brothers Revised Edition 1944. Chapt. 11.

CONCEPT FORMATION OF NORMAL AND SUBNORMAL ADOLESCENTS ON A MODIFICATION OF THE WEIGL-GOLDSTEIN-SCHEERER COLOR FORM SORTING TEST

ARNE KORSTVEDT, CHALMERS L. STACEY, AND WILLIAM F. REYNOLDS

*Syracuse University**

PROBLEM

In 1937, Bolles[1] reported on a study of concept formation in which she used three groups of subjects: ten adult aments, nine adult dements, and ten normal children. The test used by Bolles was the Weigl Form-Color Sorting Test. She concluded that the aments tended to sort the figures on the basis of form, and also showed an inability to shift attitude voluntarily and sort the figures according to another concept.

The major drawback of the Bolles study was the very small number of subjects included in the experiment. The major purpose of the present research was to study the concept formation of normal and subnormal subjects, using a relatively large number of subjects, on a modification of the Weigl-Goldstein-Scheerer Color Form Sorting Test.

METHOD

The subjects for the experiment were 50 normal and 75 subnormal females, ranging in age from 15 to 18 years. The normal girls had a mean IQ of 109. The subnormal subjects were divided into a moron and a borderline group. The 45 moron females (48-69 IQ) had a mean IQ of 60. The 30 borderline subjects (70-79 IQ) had a mean IQ of 74. The IQ's of the subnormal subjects had been determined by the use of the Revised Stanford-Binet, Form L. The IQ's of the normal subjects had been determined by the use of the Otis Self-Administering Test of Intelligence.

The test employed in the present study was a modification of the Weigl-Goldstein-Scheerer Color Form Sorting Test[2]. The materials included 24 plastic figures: six squares, six equilateral triangles, six acute right triangles, and six four-sided parallelograms (diamonds). One figure of each form was in each of the following colors: red, yellow, green, blue, black, and white. All figures were white on the reverse side. The concept formation test was administered individually to every subject.

The testing procedure was as follows:

1. The figures were placed before the subject in a random order, colored side up. The subject was asked to group the figures that seemed to belong together, and to state the principle of his performance.

2. The subject was asked to sort the figures in a way different than the way he had previously, and to state the basis of the grouping.

3. For those subjects who had grouped the figures by both form and color on trials 1 and 2, the experiment was over. Some subjects grouped by only one category on the first two trials. Other subjects "patterned" the figures in some form of design. For any subjects not successful in sorting the figures by both form and color on the first two trials, the following procedures were followed:

(a) If the subject grouped only by form on trials 1 and 2, the experimenter sorted the figures by color and explained the basis of the grouping. Then, the first two steps of the procedure were repeated, but the trials were called trials 3 and 4.

(b) If the subject had sorted only by color on trials 1 and 2, the figures were turned white side up, eliminating the color factor. If the subject was still unable to sort by form, the experimenter demonstrated and explained the grouping. The subject was then given trials 3 and 4.

---

*The authors wish to express their gratitude to Dr. S. W. Bisgrove, Senior Director, Syracuse State School, and to Miss Catherine Heffernan, Guidance Counselor, Marcellus Central School, for their cooperation in this study.

Results

From Table 1, it can be seen that, in all but one comparison, significant differences were obtained between the normal and subnormal subjects in their performance on the concept formation test. The normal females were much more successful in sorting the figures by both color and form than were the subnormal subjects. This result is in harmony with the results of the Bolles study. Table 1 also reveals that there were no significant differences between the borderline and moron subjects. Both groups had equal difficulty in sorting by two concepts, in that less than 25 per cent of both groups sorted by both color and form on the first two trials.

Table 1. Sorting Performance of All Subjects on the First Two Trials

Sorting Performance	Normals N=50	Subnormals N=75	Critical Ratio	Morons N=45	Borderlines N=30	Critical Ratio
1. Sorted by Both Form and Color	41	18	5.11**	11	7	.81
2. Sorted by One Concept only	3	17	2.54*	7	10	1.70
(a) Sorted by Color only	2	7	1.56	3	4	.20
(b) Sorted by Form only	1	10	2.24*	4	6	.52
3. Sorted by Neither Concept (Patterned)	6	40	6.54**	27	13	1.42
4. Sorted First by Color	20	13	2.87**	9	4	1.36
5. Sorted First by Form	24	22	2.70**	9	13	.82
6. Stated Basis of Sorting (Color 1+2a)	43	25	5.88**	14	11	.60
7. Stated Basis of Sorting (Form 1+2b)	42	28	5.11**	15	13	.91
8. Sorted by Form only after White Side Turned up (2a+3)	8	18	3.40**	11	6	.13

*P is less then .05.
**P is less than .01.

A much higher percentage of subnormals than normals were only capable of sorting by one concept on trials 1 and 2. Over one half of the total group of subnormals patterned the figures on the first two trials. The moron group tended somewhat more to this unsuccessful response than the borderline group, but the difference was not significant. Only 6 normal subjects could not group by either concept on trials 1 and 2. Eight normal subjects, unable to sort by form on trials 1 and 2, grasped the concept when the figures were turned white side up. Of the 47 subnormals who were unable to sort by form, only 18 of them were able to when the figures were turned white side up.

It can be seen from Table 1, that the subnormal subjects tended strongly to sort the figures first by form. The normal subjects showed a slight, insignificant tendency to sort first by form. The subnormal group's preference for form was accounted for by the borderline group, which sorted first by form by more than three to one. The moron group showed no such preference, sorting first by each concept an equal number of times. This result is in direct opposition to the findings of Bolles that aments tend to sort first by color.

It can be seen from the data in Table 2, that 9 of the 50 normal subjects were required to take trials 3 and 4. Every one of them was able to group the figures by

TABLE 2. SORTING PERFORMANCE OF NORMAL, BORDERLINE, AND MORON SUBJECTS ON TRIALS 3 AND 4

IQ Group	Number of Subjects Given Trials 3 and 4	Subjects Who Sorted By Both Form and Color N	%	Critical Ratio
Normal	9	9	100	2.00*
Borderline	23	16	70	.17
Moron	34	23	68	

*P is less than .05.

both form and color following the demonstrations. Of the 23 borderline subjects who took trials 3 and 4, 16 of them profited from the demonstrations and sorted successfully by both concepts. Of the 34 moron subjects who took trials 3 and 4, 23 of them were able to group the figures by both color and form. It is of interest to note that approximately the same percentage of moron and borderline subjects succeeded after the demonstrations. Approximately 30 per cent of each group failed on trials 3 and 4. The difference between the normal and subnormal groups on trials 3 and 4 was significant. There was no difference between the moron and borderline groups on trials 3 and 4.

SUMMARY

A modification of the Weigl-Goldstein-Scheerer Color Form Sorting Test was administered to a group of 50 normal and 75 subnormal adolescent females. The subnormal group was composed of 45 morons (48-69 IQ), and 30 borderlines (70-79 IQ). The materials included 24 plastic figures in four designs and six colors. Each subject was given two trials to sort the figures by color and form. Subjects who were not successful in grouping by both concepts had the sortings demonstrated to them, and then were given trials 3 and 4.

In all but one statistical comparison between the normal and sub-groups on trials 1 and 2, large and significant differences were found. The subjects with normal intelligence were much more successful in sorting the figures by both color and form. All statistical comparisons between the moron and borderline subjects were insignificant, indicating that both subnormal groups had about equal difficulty on the concept formation test. Borderline subjects showed a decided tendency to sort first by form, while normal and moron subjects showed no preference. All normal subjects who took trials 3 and 4 were successful in sorting by both color and form. In both the moron and borderline groups, approximately 70 per cent of the subjects who took trials 3 and 4 were able to sort by both concepts. The difference between the normal and subnormal groups was statistically significant.

REFERENCES

1. BOLLES, MARY M. The Basis of Pertinence. *Archives of Psychology*, 1937, No. 212.
2. GOLDSTEIN, KURT and SCHEERER, MARTIN. Abstract and Concrete Behavior: An Experimental Study with Special Tests. *Psychological Monographs*. 1941, 53, No. 239.

HIDDEN COSTS IN THE UTILIZATION OF THE PSYCHIATRICALLY MARGINAL MAN*

WILLIAM A. HUNT
Northwestern University

CECIL L. WITTSON
University of Nebraska College of Medicine

and

EDNA B. HUNT

Introduction

In evaluating the utilization by the military services of the psychiatrically marginal individual, the major attention centers around whether or not the individual concerned can successfully complete a stated period of military duty without the occurrence of a psychiatric breakdown which necessitates medical discharge from the service. Little attention is paid to the quality of the service rendered. In a previous article[3] the present authors studied the hospital and disciplinary records of a group of known psychiatrically marginal individuals who completed a three-year period of military duty terminating in honorable discharge. The incidence of hospitalization and disciplinary difficulty for this marginal group, however, was much greater than that found for a control group of normal individuals. Because of the importance of these findings for military manpower policy and planning, it was decided to repeat the investigation using a new group of subjects.

Procedure

Both the experimental and control groups used in the present study consisted of recruits taking their basic training at the U. S. Naval Training Center, Newport, R. I. during the year 1943 who were referred to the Psychiatric Unit for observation because of suspected neuropsychiatric difficulty and who fulfilled the following criteria: they must have been studied by at least two staff members with unanimous agreement reached, either as to psychiatric diagnosis or the absence of positive findings ("normal"); they must have been adjudged capable of service and returned to full duty; they must subsequently have served at least three years of active duty terminating in honorable discharge from the service. The normal control group contained 79 subjects. The experimental groups consisted of 65 neurotics of varying type, 31 schizoid personalities, 28 alcoholics, 69 individuals of low intelligence (mental age below 12 years), and 45 asocial psychopaths. These men represent every available case under these diagnoses that could be found in the Unit's records for 1943 and that fit the somewhat exacting criteria stated above and whose subsequent service records were available at the time of the study. It should be stressed that the original diagnostic findings were entered only in the Unit's professional records and not on the recruit's permanent health record where it might prejudice his subsequent career in the Navy.

The hospital and disciplinary histories of these men were then followed until December 31, 1945, an average period of two and one-half years. The analysis of hospitalization and disciplinary difficulty was made directly from the medical and service records of each man. These were obtained from the Naval Records Management Center, Garden City, Long Island. Some cases originally intended to be included in the sample were lost owing to the unavailability of service records, but checks introduced in previous studies of ours have shown that this does not entail any systematic bias[2, 3]. Hospitalization was defined as any admission to a Naval hospital for any reason whatsoever, and disciplinary difficulty was defined as any court-martial, ranging in severity from a Deck Court to a General Court-martial.

*This study is part of a larger project subsidized by the Office of Naval Research under contract 7onr-45011 with Northwestern University. Thanks are due Mrs. Jean Moore, and the staff of the Naval Records Management Center at Garden City, Long Island, for assistance in obtaining the medical and service records involved. The opinions expressed are those of the individual authors and do not represent the opinions or policy of the Naval service.

Results

The results are presented in Table 1, which records the percentage of each group who were hospitalized or in disciplinary difficulty one or more times. Previous experience has shown that this simple percentage is an adequate measure of group differences, and agrees closely with other measures that might have been used such as total number of incidents involved, time lost from duty, etc.[4] The finding of the original study that psychiatrically marginal individuals who successfully complete a term of Naval service nevertheless show a higher incidence of hospitalization than do a group of normal controls is confirmed. Inspection of Table 1 shows that the differences between the marginal groups and the normal group are all in the expected direction. With the exception of the group of psychopaths, all the differences are statistically significant at the 5% level.

Table 1. Hospitalization and Disciplinary Action in Normal and Marginal Groups

Classification	N	% Hospitalized	% Disciplinary Cases
Normal controls	79	22	23
Neurotics	65	46	14
Schizoid personalities	31	42	30
Alcoholics	28	39	36
Low intelligence	69	43	36
Psychopaths	45	30	53

The disciplinary data also confirm the original study in each instance. With the exception of the neurotic group, all the marginal groups show more disciplinary difficulty than does the normal control group. Only for the low intelligence and psychopathic groups, however, were the differences significant at the 5% level. The difference between normals and neurotics approaches the 8% level.

These results should not be interpreted as indicating that the marginally adjusted man cannot be utilized by the military services. They merely show that such men are more "expensive" to the services, and that when their utilization is demanded by manpower needs it will be necessary to make provision for the added demands they will entail upon medical and disciplinary facilities. They also confirm the clinical picture of the maladjusted individual as one who, even when he is meeting the formal adjustmental standards of his group, is doing so at a greater cost to the group's medical facilities and with greater friction upon its social organization than is his adjusted compatriot.

Summary

A comparison of a group of marginally adjusted individuals and a control group of normals, both of whom successfully completed a three-year term of Naval service shows that despite their successful completion of service the marginal group had a higher incidence of hospitalization and of disciplinary difficulty. The added demands that such marginal groups make upon hospital and disciplinary facilities in the military services must be considered in manpower planning.

References

1. Hunt, W. A., Wittson, C. L. and Burton, H. W. A validation study of naval neuropsychiatric screening. *J. consult. Psychol.*, 1950, 14, 35-39.
2. Hunt, W. A., Wittson, C. L. and Burton, H. W. A further validation of naval neuropsychiatric screening. *J. consult. Psychol.*, 1950, 14, 485-488.
3. Hunt, W. A., Wittson, C. L., and Hunt, E. B. Military performance of a group of marginal neuropsychiatric cases. *Amer. J. Psychiat.*, 1952, 109, 168-171.
4. Wittson, C. L., Hunt, W. A., and Stevenson, I. A follow-up study of neuropsychiatric screening. *J. abnorm. soc. Psychol.*, 1946, 41, 79-82.

CORRELATES OF THE NON-COMMITTAL TEST-ITEM RESPONSE

GRACE RUBIN-RABSON, PH.D.

1001 E. Rudisill Blvd.
Fort Wayne, Indiana

Purpose

The omission of an item on an intelligence test may indicate frank ignorance of the response or, in the speed situation, a judgmental evaluation of best procedures. Omitted items are then an intrinsic contribution to the intelligence score. On an attitude scale, however, where a zero response is utilized to indicate uncertainty, lack of decision, lack of information or conviction, ignorance may be only one ingredient with some participation of other personality factors as well.

It is proposed to examine here the correlates of this zero or non-committal response, which, like the "don't know" response, may be, according to Marks[8] an evasion in a stigmatized situation.

Previous Studies

Since its inception the Bernreuter Personality Inventory has been a widely used and easily administered instrument for the analysis of personality in its more superficial aspects. In 1946 Ellis[3] observed the paucity of studies showing objective or clinical validation of the test and instead an emphasis on statistical checks and balances.

That the inventory differentiates marked deviates from normals under certain conditions seems clear[14], but except for the B 4D scale's distinction of normals and psychoneurotics, the scores of the other three scales are not significantly different. Furthermore, the scores can be slanted at will. Kimber[6] was able to produce good adjustment scores by specific instructions toward this end. Sartain[13] found the reliability of scales low, ranging from .68 to .52. Martin,[9] using Bennett's simplified scoring stencils, traced most of the inventory's effectiveness to the intro-extroversion and dominance-submission scales. Lorge[7] found an interdependence of all total trait scores, nor was any trait sufficiently discrete for clinical diagnosis.

Seeking other correlates of the BPI, Sanai and Pickard[12] in 1946 found that Rorschach patterns do not support any of the popular, psychoanalytic or clinical assumptions regarding the relationships between radicalism and "temperamental traits." Dexter[2] in 1938 presented correlations of radicalism and three BPI traits between .24 and .28.

In his summary, Kerr[5] finds the liberal more introverted and intelligent. Eysenck,[4] in a British sample, demonstrated extroversion in liberals and those preferring the modern school of art.

Nor do these personality traits depend on intelligence. Ellis seems to review no studies in this area[3]. Zalman[15] comparing BPI results and academic success, disclosed one significant "r" of .23 between introversion and military science, whatever this may signify about intelligence.

Procedure

In search of conservative-liberal attitude correlations, a group of 43 adults,* 18 to 58 years of age, attending a series of lectures in psychology at the Indiana University Extension in Fort Wayne, Indiana, during the fall of 1949, provided data on the Hunter Test of Social Attitudes, the Bernreuter Personality Inventory, the verbal section of the Bellevue-Wechsler Intelligence Test and several personal items such as schooling, age, number of siblings, etc. The majority were office and business persons, the rest professionals, executives and a few factory workers.

For the group, mean age was 31 (S. D. 10.63), mean formal schooling, 13 years, and relatively homogeneous in this respect (S. D., 1.61). An intelligence average of I. Q. 114 (S. D. 10.50) placed the group somewhat above the average though allowance might be made for some possible overestimation by the Bellevue-Wechsler, as

*This group, part of a larger group, remained anonymous in the several testing situations of the experiment.

revealed in a previous study[11]. Assuming the comparability of the experimental and the norm group on three of the BPI trait measures, the 43 subjects seemed normally introverted (57th percentile, S. D. 29.67); normally self-sufficient (47th percentile, S. D. 23.74); but more than averagely submissive, (36th percentile, S. D. 26.60). According to Adorno et al[1] submissive groups tend toward the authoritarian characteristics of the fascist mentality with its concomitant racial biases. This group, though somewhat submissive and more than a little conservative in five of seven sections of Hunter's test, showed unusual liberalism toward the Negro[11].

Results

Table 1. Correlates of Several Variables and Number of Non-Committal Responses

Variable	Non-Commital Responses*	Intelligence	Liberal Conservative Attitude Score
Introversion-extroversion	.14	.09	.000
Dominance-submission	−.26	−.07	.04
Self-Sufficiency	−.43	.04	.09
Age	−.05		
Schooling	−.07		
Intelligence	−.31		.40

*The correlations of non-committal responses, age and schooling derive from a population of 69; of the same and intelligence, from a population of 49; of the same and Bernreuter scores, 43.

Of a large series of correlations, most of them reported elsewhere[11], Table 1 presents only those necessary to highlight an unexpected finding. Of the relationship of conservative-liberal attitudes and the Bernreuter scale the evidence is conclusively negative. Though attitude is dependent on intelligence it hardly seems to relate to personality characteristics as revealed by the BPI.

Among five answer categories, the Hunter test includes the zero response for uncertainty, lack of decision, lack of information or conviction. If lack of decision and conviction are attributable to lack of information, then, according to an intelligence test such as the Bellevue-Wechsler which includes information as a sub-test, there should be some relationship between number of responses marked 0 and intelligence. This proves to be the case. The negative correlation, −.31, is significant at better than the 5% level and might be higher in a more heterogeneous population. (The B-W IQ ranged in this group from 98 to 138.) However, though intelligence as measured by the B-W seems an influential factor, the information sub-test itself yields a less than significant correlation of −.24 with number of responses marked 0.

Another hypothesis suggests that the more intelligent might resort to the non-committal response as a rejection of the idea or the wording of the test statement. This would operate in the direction of a positive correlation, serving to mask other factors making for a negative correlation.

If the information and intelligence factors contribute in a measure to this degree of dependence, it must be concluded nonetheless that schooling plays little part in producing this variable since the r of schooling and number of responses marked 0 is a negligible −.07. Nor does the presumed accretion of information with age reveal any operation in the results.

In addition to the information and intelligence ingredients is the uncertainty and lack of decision involved, and this shows a fair measure of dependence on the degree of individual self-sufficiency. The correlation of −.43, though not high, is significant at better than the 1% level, and possibly implies an intrinsic relationship worth further investigation.

During historically uncertain periods, the avoidance of attitude test statements might indicate simple unwillingness to commit oneself on suspect topics. The "don't know" response would then, according to Marks[8], be an evasion in a stigmatized situation. But since S in the present experimental frame remained anonymous, this possibility is untenable except as "evasion" is a general personality characteristic regardless of the stigma in the situation.

A feeling of self-insufficiency is a basic dynamic in the avoidance pattern, whether of persons, activities or ideas. Any test, therefore, which includes an option on evasion might be expected to yield, in addition to its primary function, this extra clinical diagnosis.

As Rabin and Guertin[10] observe in another connection, "almost any test employed by a skilful clinician or insightful worker may prove an 'aid in diagnosis'." The omitted or non-committal response, whether on an intelligence test, an attitude scale or on one more directly designed for personality analysis, is recommended to the insightful worker as worthy of consideration.

Summary

In a series of variables concerned with personality traits, only two showed significant correlations: The number of non-committal responses on a conservative-liberal attitudes scale with intelligence, and the same with degree of self-sufficiency. Neither age nor length of schooling played any considerable part nor did the information sub-test of the Bellevue-Wechsler Intelligence Test yield a significant relationship with the number of responses marked 0 despite the fact that information is presumably operative in determining opinion.

More clearly manifest is the dependence of lack of certainty, decision and confidence on the degree of feelings of self-insufficiency as measured by the Bernreuter Personality Inventory.

The results of any test allowing the omission of items or the choice of a non-committal response might then be appraised in terms of this additional personality facet.

Bibliography

1. Adorno, T. W., Frenkel-Brunswick, E., et al. *The Authoritarian Personality.* New York: Harper, 1950.
2. Dexter, E. Personality traits related to conservatism and radicalism. *Charac. and Personal.*, 1938, *7*, 230-237.
3. Ellis, A. The validity of personality questionnaires. *Psychol. Bull.*, 1946, *43*, 385-440.
4. Eysenck, H. J. Personality factors and preference judgments. *Nature*, 1941, *148*, 346. London.
5. Kerr, W. A. Correlates of politico-economic liberalism-conservatism. *J. soc. Psychol.*, 1944, *20*, 61-77.
6. Kimber, J. A. M. The insight of college students into the items on a personality test. *Educ. Psychol. Measmt.*, 1947, *7*, 411-420.
7. Lorge, I. Personality traits by fiat. 1. The analysis of the total trait scores and keys of the Bernreuter Personality Inventory. *J. educ. Psychol.*, 1935, *26*, 273-278.
8. Marks, E. S. The undecided voter. In Mosteller, F. et. al. The pre-election polls of 1948, 263-289. Report to the Committee on Analysis of Pre-Election Polls and Forecasts. *N. Y. Soc. Sci. Res. Council*, 1949.
9. Martin, G. C. A factor analysis of the Bernreuter Personality Inventory. *Educ. Psychol. Measmt.*, 1948, *8*, 85-92 .
10. Rabin, A. I. and Guertin, W. H. Research with the Wechsler-Bellevue Test. *Psychol. Bull.*, 1951, *48*, 211-248.
11. Rubin-Rabson, G. Several correlates of a conservatism-liberalism attitudes scale. *J. soc. Psychol.*, in press.
12. Sanai, M. and Pickard, P. M. The relation between politico-economic radicalism and certain traits of personality, *J. soc. Psychol.*, 1949, *30*, 217-227.
13. Sartain, A. Q. A study of Bernreuter Personality Inventory scores made by candidates for supervisory positions in an aircraft factory. *J. soc. Psychol.*, 1946, *24*, 255-259.
14. Schmidt, H. O. and Billingslea, F. Y. Test profiles as a diagnostic aid: The Bernreuter Inventory. *J. abnorm. soc. Psychol.*, 1945, *40*, 70-76.
15. Zalman, W. R. The relationship of the traits of the Bernreuter Personality Inventory to academic success. *J. Amer. Ass. colleg. Registr.*, 1945, *21*, 81-84.

THE RELATIONSHIP BETWEEN Z SCORES ON THE BENDER GESTALT AND F+% ON THE RORSCHACH*

ROBERT H. CURNUTT, LT. (JG) MSC, USNR AND WILLIAM B. LEWIS, ENS. MSC., USNR

*U. S. Naval Hospital, National Naval Medical Center**

Introduction

Recently a scoring system for the Bender Gestalt test has been developed by Pascal and Suttell [3]. By tabulating certain deviations in the reproductions of the designs a raw score is obtained which in turn is converted into a Z score. These authors believe that the higher the Z score the greater is the likelihood that we are dealing with a psychiatric patient, while the lower the Z score the more integrated is the personality adjustment of the individual. This integrative capacity of the organism may be thought of as a function of the ego, for we find them saying: "If, for the purposes of this discussion, we limit our definition of ego function to the ability to reproduce faithfully the B-G drawings as presented, then we may say that ego strength lies on a continuum from very low to very high B-G scores. This quantitative continuum may, then, in a very limited way, be thought of as an aspect of ego strength—the lower the score the greater the ego strength." [3]

Beck on the other hand believes that F+ on the Rorschach is also a measure of ego strength. He says: "Applying this criterion in the several personality groups studied by means of the Rorschach test, the striking fact is that the weaker the ego, the more F− percepts appear. The converse also holds: the stronger the ego, the more F+ associations." [1]

At first glance it appears that Pascal and Suttell claim to be measuring the same function with their Z score that Beck believes the F+% measures. One would expect to find an inverse relationship between these scores. That is, in a given population, the higher the F+%, the lower will be the Z score. It should be recognized, however, at the outset, that these authors may be talking only about certain specific aspects of "ego strength," or they may merely be using similar language in describing unrelated factors. If these two "ego strengths" or "aspects of ego function" are essentially the same there should be a high negative correlation between these two sets of scores.

The hypothesis to be tested may be stated as follows: The Bender Gestalt Z score measures primarily the same function (ego-strength) as does the Rorschach F+%.

Method

The records of 25 patients who had been hospitalized in a Naval Neuropsychiatric Unit were withdrawn from the files. The only bias in their selection was that they had all been given the Rorschach and Bender tests by the same experienced clinical psychologist. This was done to insure that any individual deviation in the F+% scoring would be constant. The Bender records were scored according to the method of Pascal and Suttell. The author has found that his scoring average is 3.6 points above that of Pascal and Suttell who differ by 4 or 5 points between themselves.

Our sample consisted of 7 patients with organic brain damage, 8 schizophrenics, 9 neurotics, and one case of psychosomatic asthma. F+% and Z scores were correlated using the Rank Difference Method.

Results and Discussion

A Rank Difference coefficient (Rho) of .07 was obtained for the Z scores and F+% of the group. Thus the expected relationship between F+% and Z scores was not realized. It appears then that they do not measure the same function in this mixed clinical group. Perhaps the F+% does not measure ego strength while the Z score does measure it, or the converse may be true. Again, the two scores may be measuring different, mathematically unrelated aspects of ego strength.

---

*The opinions contained herein are the private ones of the writers and are not to be construed as official or reflecting the views of the Navy Department or the Naval Service at large.

A third possibility is that these scores measure different functions in discrete nosological groups, or even in different individuals.[1] The authors favor this alternative, reasoning as follows: It is falacious to assume that a given Z score or F+% can be explained in terms of some constant factor which the Z or F+% always measures. What is measured is a function not only "of the measuring instrument" but of what there is to be measured. The Bender Gestalt Z score and the Rorschach F+% vary in relation to an unknown number of factors. One cannot assume that the primary variable in a specific case is "ego strength" (or anything else) in the absence of further knowledge of the individual tested.

Summary

In order to determine the relationship between two reputed measures of ego strength, the Bender Gestalt Z score and the Rorschach F+%, these scores from a heterogeneous group of 25 neuropsychiatric patients were correlated, using the Rank Difference method. No significant relationship was found. The implications of this finding are discussed.

References

1. Beck, S. *Rorschach's Test II. A Variety of Personality Pictures.* New York: Grune and Stratton, 1947.
2. Curnutt, R. H. The Use of the Bender Gestalt With An Alcoholic and Non-Alcoholic Population. *J. clin. Psychol.* 1953, 9, 287-290.
3. Pascal, G. and Suttell, B. *The Bender-Gestalt Test.* New York: Grune and Stratton, 1951.

[1]The Z scores and F+% of the 7 brain-damaged patients in the sample, considered alone, yielded a Rho of −.58. Though not statistically significant for so small a sample, it may indicate a trend within this nosological group.

CORRELATION OF CLINICAL ESTIMATES WITH TEST SCORES ON MENTAL ABILITY AND PERSONALITY TESTS

J. WATSON WILSON

Nordli, Ogan, Wilson Associates
New York, N. Y.

Problem

This study is concerned with the predictability of performance on intelligence and personality scales through the use of a structured, clinical interview. It raises, and perhaps partially answers, the question: with what degree of accuracy can a clinician, using a history-taking procedure, predict performance on such tests?

Data and Methods

With no particular research in view, the writer has long attempted to estimate the client's test scores prior to the administration of the tests. An examination of case files of four client companies (a large health and accident company, a small fire insurance company, a large friction materials manufacturer, and a large corporation manufacturing fuel injection and electrical equipment) yielded 250 cases on which estimates were recorded for several different tests. Three tests were found to have been used with sufficient frequency over a two year period to warrant computing correlations. A wide variety of personnel were included in the group studied: engineers, salesmen, accountants, underwriters, officers, and foremen.

Before discussing the data, it would be advisable to provide a context for the predictions. With each individual, the clinician took a detailed personal history. It consisted of job, educational, family, social, and health histories. This typically took from 20 to 40 minutes.

In taking the history certain general hypotheses with respect to the individual were established. At this point a personality inventory such as the Personality Record (a short form adjustment inventory) was administered. While the evaluee was completing the personality inventory, significant parts of his history were isolated for further study and highly individual questions were prepared for a deeper exploration. From this study came further hypotheses and suggestions for testing. The Beckman, oral projection or parts of the T. A. T. frequently were used at this point.

A mental ability test such as the Personnel Problems Test, Otis, Wonderlic, or Wechsler-Bellevue was administered. In each instance, predictions were recorded before the tests were given.

Results

On the Personality Record, scores (atypical responses) and predictions were available for 75 cases. Figure 1 is the scattergram for these comparisons. The atypical responses ranged from 2 to 24, a score of 2 being at the "well adjusted" end of the scale and a score of 24 being extreme on the "neurotic" end. The correlation between predictions and scores was .81 (σr_{xy} = .038). The average scale unit differences between the predictions and the score was 1.5.

The psychologist underestimated test scores in 25 cases (33.3 per cent), and overestimated in 31 cases (41.3 per cent).

On the Beckman, predictions were available for 70 cases. Here, predictions merely indicated the group or decile in which the score would fall, i. e. A4, A3 — to — S3, S4. The range was from S4 (the extreme on the submissive end) to A4 (the extreme on the aggressive end). The correlation between predictions and scores was .66 (σr_{xy} = .067). The average of the differences was 1.1. The scattergram for this comparison is contained in Figure 2. In 22 cases (31 per cent) the psychologist underestimated aggressiveness and in 23 cases (33 per cent) he overestimated it.

On the Personnel Problems Test (a short-form, intelligence test) it was possible to calculate correlations between predictions and scores with respect to both items tried and items answered correctly. Figure 3 is the scattergram for the comparison of predictions and scores with respect to items answered correctly. Data were available on 160 cases. Here the range of scores was from 11 (the 17th percentile) to 34 (the 99th percentile). The correlation between predictions and scores was .74 (σr_{xy} = .036). The average of the differences was 2.0. In 74 cases (46 per cent) the psychologist overestimated the mental ability score while he underestimated it in only 54 cases (34 per cent).

Figure 4 is the scattergram for the comparison of predictions and results with respect to the number of items the individual would attempt. Here only 149 cases

Figure 1

Relationship of Predictions to Scores on the "Personality Record"

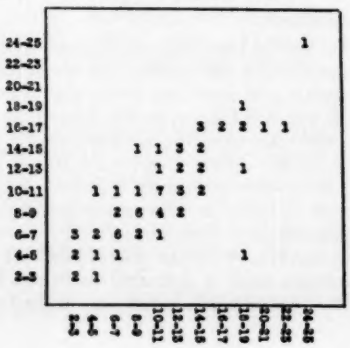


Figure 2

Relationship of Predictions to Scores on the "Beckman"

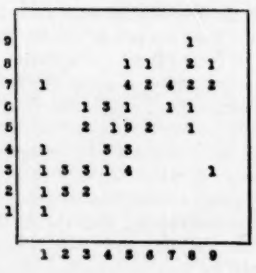


Figure 3

Relationship of Predictions to Correct Answers on the "Personnel Problems Test"

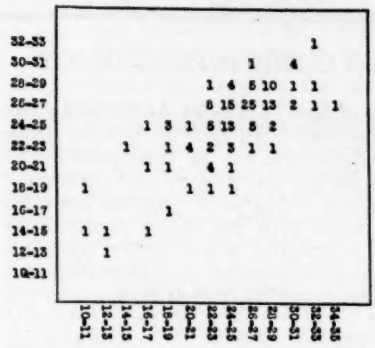
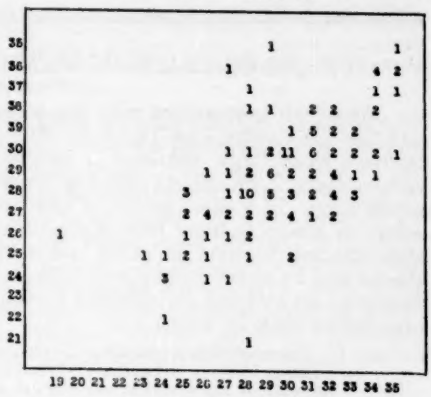


Figure 4

Relationship of Predictions to the Number Attempted on the "Personnel Problems Test"



were available. The range was from 19 (the 57th percentile) to 34 (the 99th percentile). The correlation between the prediction regarding the number of items the individuals would try and the number they actually did try was .65 (σr_{xy} = .047). The average of differences was 1.7. Whereas the psychologist tended to overestimate the individuals' mental ability, he tended to underestimate the number of items the individuals would attempt. He underestimated in 67 (45 per cent) and overestimated in only 35 cases (23.5 per cent).

Suggestions and Questions

Research of this sort may raise more questions than it answers. However, it offers certain suggestions:

1. Following a patterned interview with which he is familiar, using tests with which he is familiar, and administering psychometric instruments to test specific hypotheses, the clinician can predict with a fairly high degree of accuracy how the individual will perform or "score". Actually, the correlation between the clinician's prediction and the "score" is somewhat higher in certain instances than correlations between supposedly comparable tests. This, of course, raises the question, whether or not the clinician should rely more on his clinical judgment and less on test scores.

2. A carefully structured interview may have greater validity with respect to predicting level of ability, personality structure and the like than historic studies of the interview would indicate. This is counter to historic studies which suggest that characteristics such as those examined here cannot be adequately discerned on the basis of an interview.

3. Finally, the question must be raised whether the findings here can be duplicated by others. While this may be problematical, the author holds no brief either for his own techniques or his particular skill in the use of them. Certainly other studies of this sort are needed. Such studies, adequately planned, could include a wider variety of tests than was possible here.

EDITORIAL OPINION

CRITIQUE OF TRAINING PROGRAMS IN CLINICAL PSYCHOLOGY

American universities with the cooperation of the VETERANS ADMINISTRATION and the EDUCATION AND TRAINING BOARD of the AMERICAN PSYCHOLOGICAL ASSOCIATION have truly achieved a prodigious accomplishment in establishing and standardizing high calibre training programs in clinical psychology and counseling psychology. In a short period of less then 10 years, a degree of progress has been achieved which is truly remarkable when compared with the slower evolution of standardized training in other specialties. The remarks which follow therefore should not be construed as involving any blanket criticism or condemnation, but simply as an attempt to highlight certain features in which further progress in the near future may be hoped for.

1. *Too much didactic teaching.* There appears to be too much emphasis on organized formal courses, lectures, symposia and theoretical discussions. Time spent on formal course requirements far overbalances clinical ward work. Students should get their theoretical orientation from required readings, with only interpretive comments from formal lectures.
2. *Reliance upon Authority.* There seems to be a tendency to gain respectability for clinical psychology training programs from basking in the reflected light of "big names" and authorities in neighboring fields. Students spend too much time listening to theoretical discourses from these authorities who often tend to "ride" their own interests or schools.
3. *Preoccupation with Theory.* Trainees finish their programs having invested an inordinate amount of time learning about theories and relatively little time in face-to-face clinical contacts. A more operational approach would acquaint the student more organically with exactly what a clinical psychologist actually does in a wide variety of situations. Many of the hours now given to didactic lecture work could better be devoted to supervised clinical contacts.
4. *No Psychological Ward Rounds.* The most glaring defect of training programs is the lack of psychological ward rounds and clinical psychopathological conferences in which students could come face-to-face with a wide sampling of case materials under personal supervision of instructors who would point out important points in "bedside" teaching.
5. *Inadequate Experience with the Whole Person.* Before beginning the study of more specialized techniques, the student should first be placed on the wards or in the clinic, and exposed to all the little details of how to relate to patients, how to influence them for desired ends, and in general to gain an organic experience with their conditions of living. Each student should *first* become adept in face-to-face case evaluation and handling. Clinical judgment can only be developed in this manner. This experience might well be gained by starting the trainee in as a ward attendant for three month periods during summer vacations.
6. *Premature Introduction to Testing.* In many installations, the student is introduced to clinical contacts first and only through testing. This approach is very harmful in that the student gains little first hand experience in developing his own clinical judgment and diagnostic ability but is thrown into a more or less mechanical application of technical skills. Testing should come much later only after basic clinical skills are developed.
7. *Unbalanced training experiences.* There is an unfortunate trend in the direction of overemphasing the psychoneuroses and psychoses at the expense of other areas such as mental deficiency, child guidance, delinquency, marriage counseling and other problems of equal social importance. With respect to mental deficiency, for example, this is a field to which clinical psychologists can make a great contribution and also in which there is great social need. All training programs should include at least 6 months in an institution for mental defectives as well as representative clerkships in other important areas. It is not unusual to find graduates of current training programs who have had adequate experience only with neurotics and psychotics. This reflects the undue influence of mental hospitals in their training.
8. *Standards too Inflexible.* While it is desirable for standards established by the Veterans Administration and the APA boards to be sufficiently rigid, there is also the undesirable possibility that extreme rigidity may inhibit flexibility in working toward better programs. It is understandable that the Veterans Administration should desire that training programs provide the necessary background for VA work, but too rigid an application of such policies will leave the trainee unbalanced for other responsibilities which he may be called to assume outside VA auspices. The VA and the APA should permit recognized university

departments to experiment with the broadening of training programs. The present situation is comparable to that which would occur in medical education if the Army supervised all training in the direction of its own needs.

9. *Research vs Clinical Proficiency.* To our knowledge, it has never been scientifically proven that the good researcher is always the good clinician, or vice versa. On the contrary, clinical experience suggests the opposite. Research ability is a rare quality which should not be equated with clinical sensitivity. Clinically sensitive trainees should not be held back if they happen to lack research qualifications.

In any program for evaluating clinical training facilities, there is a tendency for the "seal of approval" to gain almost a magical or mystic significance, constituting as it does the mark of approval of existing "authorities". The agency seeking approval may even tend to sacrifice its own unique potentialities for growth because of the need to submit to whatever standards may be in vogue. It is not beyond the borders of possibility to conceive of a situation where some "school" of thought might gain sufficient ascendancy so as to be able to dictate nationwide standards based on its own particular pattern. The Behaviorists succeeded in dominating the American scene in this manner for a few years, just as psychoanalytic influences have gained overwhelming influence in some parts of the contemporary psychiatric scene. Let us hope that hope that standards in clinical psychology remain flexible and eclectic in their orientation.

F. C. T.

LET'S BURY THE DEAD

A visit to any drugstore will provide convincing evidence of the truism that obsolete medicines continue in popular use far after they have been proven worthless and supplanted by more valid medications. This is true not only of patent medicines which are kept alive by their producers for commercial reasons, but also of secondary medicants which may have some therapeutic value but which have been replaced by more potent or less toxic substances. The medical profession has attempted to differentiate between substances of differential usefulness by establishing three classifications of drugs according to their usefulness and proven validity. *The U. S. Pharmacopeia* provides a general listing of some nine or ten thousand substances for which some therapeutic value is known or claimed. It includes many substances which are now considered obsolete or have been replaced by drugs of greater potency. *New and Nonofficial Remedies* presents a compilation of claims and facts concerning new drugs or proprietary remedies which have only recently appeared with insufficient evidence to establish their potency and specificity. Finally, the American Medical Association publishes a small book *Useful Drugs* which lists the few hundred substances which have been clearly established to have *specific* value in controlling the causes and symptoms of disease. In *Useful Drugs*, such substances are listed singly in order to emphasize their rational usage for specific indications. In general, medical students are not required to familiarize themselves with all the drugs in the *U. S. Pharmacopeia*, or even in *New and Nonofficial Remedies*, but are expected to concentrate on the specifics listed in *Useful Drugs*.

The medical profession is thus kept informed concerning the relatively small number of drugs having high specificity, high potency and low toxicity. Needless to say, such classifications are subject to constant modification as new developments render old practices obsolete. In a field such as medicine where almost infinite complexity is made possible by all the permutations and combinations in which drugs can be prescribed, it is very desirable to have some system to separate the chaff from the wheat so as to permit the average practitioner to discriminate what is validly useful from all the remedies which are offered him.

Clinical psychology is rapidly reaching a similar stage of development in which hundreds of diagnostic and therapeutic methods are available to choose from. Many of these methods are outdated, having been proven by research results to be invalid or inferior to improved methods. However, such instruments continue to be used both in research involving faulty design and in clinical practice where wholly invalid predictions continue to be drawn. Poorly trained clinicians throughout the country are still making socially important decisions based on the results of such tests. As an editor, we reject each year scores of papers submitted by earnest researchers who spend hundreds of hours on research which is handicapped right from the start by use of obsolete methods.

In order to render the *coup de grace* to methods which have outlived their usefulness, it seems necessary to have some official pronouncement to the effect that they are now considered obsolete and invalid. Ideally, such a proclamation could well be made by the discoverer of the method in a public statement to the effect that its use is no longer indicated. Since it usually requires five or more years of research to establish the generalized validity of claims, it might be appropriate for the originator of new techniques to issue ten and twenty year follow-up reports summarizing the author's own experience. How valuable it would be to have a statement from John B. Watson concerning his evaluation of the evolution of Behaviorism! As far as we know, only a limited number of authors have had the courage to review the shortcomings of their discoveries or even to indicate their existence. This would almost seem to be an ethical obligation.

The other alternative might be for the AMERICAN PSYCHOLOGICAL ASSOCIATION to establish a COUNCIL OF STANDARDS AND VALIDITY authorized to issue periodic statements based on a summary of research concerning methods which are considered valid or invalid for modern conditions of practice. Commercial firms offering tests or methods to the profession might be required to submit such to the Council in order to secure approval of standardization, validity and reliability. Eventually there would accumulate a list of approved tests and methods which would serve as a guide to less trained members of the profession by making available periodic evaluations. Such a list of approved and accepted methods would also be of value to nonprofessional agencies seeking psychological services in aiding them to discriminate between professional proficiency and charlatanism. It would not impede basic research but would simply apply classificatory labels to assist in discriminating between unfounded claims and research validation.

F. C. T.

BOOK NOTICES

Anon. Ethical Standards of Psychologists. Washington, D. C.: American Psychological Association, 1953, pp. 171.

Clinical psychologists and workers in related fields should familiarize themselves with the official ethical standards governing professional psychologists.

HALPERN, FLORENCE. *A Clinical Approach to Children's Rorschachs.* New York: Grune & Stratton, 1953, pp. 270. $6.00.

The author is Assistant Professor of Clinical Psychology, New York University College of Medicine and a recognized authority on Rorschach testing. This book makes an important contribution to the testing of children through its theoretical discussions supplemented by 40 clinical records from well selected illustrative cases.

SARGENT, HELEN D. *The Insight Test.* New York: Grune & Stratton, 1953, pp. 276. $6.75.

Dr. Sargent is senior psychologist of the education department of the Menninger Foundation. The Insight Test is a verbal projective test first introduced in a monograph in 1944 and consists of a series of items or armatures presenting hypothetical social situations testing the ability of the subject to understand the motives of others. The scoring system is elaborated in great detail with illustrative case materials.

WHITAKER, CARL A. AND MALONE, THOMAS P. *The Roots of Psychotherapy.* New York: Blakiston, 1953, pp. 236. $4.50.

The authors are associated with the psychiatry department of the Emory University School of Medicine. The book consists of rather elementary theoretical essays concerning foundations, psychotherapy as process, and a brief account of some brief techniques in psychotherapy. It is difficult to understand what group this book is directed toward.

NAUMBERG, MARGARET. *Psychoneurotic Art: Its Function in Psychotherapy.* New York: Grune & Stratton, 1953, pp. 148. $6.75.

This monograph presents a very interesting study of the art productions of a 20-year-old psychoneurotic girl during 34 months of art therapy. The symbolic patterns clearly reveal the underlying conflicts which were gradually clarified in association with other materials including fantasies, dreams and writings.

Various contributors. *Drug Addiction among Adolescents.* New York: Blakiston, 1953, pp. 320.

Verbatim transcriptions of the proceedings at two conferences held on November 30, 1951 and March 13 and 14, 1952 sponsored by the Committee on Public Health Relations of the New York Academy of Medicine. The prepared papers and spontaneous group discussions contain much background material for those interested in addiction problems.

DERNER, GORDON F. *Aspects of the Psychology of Tuberculosis.* New York: Paul B. Hoeber, 1953, pp. 119. $3.50.

A Psychosomatic Medicine Monograph reporting the results of comprehensive psychological studies made at the Tuberculosis Division of the Grasslands Hospital on 32 tubercular patients. No unique tuberculous personality patterns were elicited and the study also failed to demonstrate the euphoria reported in the literature.

WITTKOWER, ERIC and RUSSELL, BRIAN. *Emotional Factors in Skin Disease.* New York: Paul B. Hoeber, 1953, pp. 214. $4.00.

A Psychosomatic Medicine Monograph presenting authoritative general discussions of the psychosomatic approach to skin disease, the study and management of skin conditions, and normative and clinical data of cases showing various abnormalities of the skin and skin sensations.

WORTIS, JOSEPH (Ed.) *Basic Problems in Psychiatry.* New York: Grune & Stratton, 1953, pp. 186. $4.50.

A distinguished group of contributors express their opinions in rather general terms on the contemporary status of psychiatry, its physiological bases, psychology and culture, the validity of mental testing, schools of psychiatry and the psychosomatic symptom. The presentation is on the level of the intelligent layman.

VON FOERSTER, HEINZ (Ed.) *Cybernetics.* New York: Josiah Macy, Jr. Foundation, 1953, pp. 184. $4.00.

Proceedings of the Ninth Conference on Circular Causal and Feedback Mechanisms in Biological and Social Systems sponsored by the Josiah Macy, Jr. Foundation.

RALLI, ELAINE P. (Ed.) *Adrenal Cortex.* Proceedings of the Fourth Conference. New York: Josiah Macy, Jr. Foundation, 1953, pp. 165. $4.00.

ABRAHAM, KARL. *Selected Papers on Psychoanalysis.* New York: Basic Books, 1953, pp. 527. $6.00. First American edition.

KARPF, FAY B. *The Psychology and Psychotherapy of Otto Rank.* New York: Philosophical Library, 1953, pp. 129. $3.00. Historical Introduction.

JUNG, C. G. *Two Essays on Analytic Psychology.* New York: Pantheon Books, 1953, pp. 329. $3.75. Volume 7 of the collected works of C. G. Jung.

FIELD, MINNA. *Patients are People.* New York: Columbia University Press, 1953, pp. 244. $3.75. A medical-social approach to prolonged illness.

FLIESS, ROBERT. *The Revival of Interest in the Dream.* New York: International Universities Press, 1953, pp. 164. $3.00. A critical evaluation of post-Freudian psychoanalytic contributions.

RIESE, WALTHER. The Conception of Disease: Its History, its Versions and its Nature. New York: Philosophical Library, 1953, pp. 120. $3.75.

ALLEN, ROBERT M. *Introduction to the Rorschach Technique.* New York: International Universities Press, 1953, pp. 126. $3.00.

An elementary manual of administration and scoring for beginning students.

PIAGET, JEAN. *The Child's Conception of Number.* New York: Humanities Press, 1953, pp. 248. $5.00.

For children of average or superior intelligence, with emotional and behavior problems:

The Department of Child Psychiatry of
The Menninger Foundation

offers

Intensive individual psychotherapy in a residential school

Outpatient psychiatric and neurologic evaluation and treatment for children up to 18 years of age is also available.

THE SOUTHARD SCHOOL

J. Cotter Hirschberg, M. D., *Director* Topeka, Kansas, Telephone 3-6494

THE HOWARD INK BLOT TEST

A NEW PROJECTIVE TECHNIQUE

PRICE: $12.50 (In U. S. and Canada; foreign postage $.50 extra)

HOWARD INK BLOT TEST MANUAL

Monograph Supplement No. 10 from July 1953 Issue

Price $2.00

Please send remittance with order.
No copies can be sent for examination or on approval.

Journal of Clinical Psychology

5 Pearl Street, Brandon, Vermont

CLINICAL STAFF

MEDICAL STAFF OF PENNSYLVANIA

Robert Devereux, M.D.
Ruth E. Duffy, M.D.
Herbert H. Herskovitz, M.D.
Robert L. Hunt, M.D.
Joseph J. Peters, M.D.
Calvin F. Settinge, M.D.
Albert S. Terzian, M.D.
Walter M. Uhler, M.D.

PSYCHOLOGICAL STAFF OF PENNSYLVANIA

Milton Brutten, Ph.D.
Michael B. Dunn, A.M.
Selma J. Ebie, A.M.
Robert G. Ferguson, A.M.
Edward L. French, Ph.D.
Marguerite B. Horn, A.M.
John R. Kleiser, Ph.D.
Kathryn Kramer
Morgan W. McKean, A.M.
Robert A. Semple, III, B.R.E.
Jack Shelley, M.Ed.
Helyn C. Wood, A.M.

PROFESSIONAL STAFF, THE DEVEREUX RANCH SCHOOL, CALIFORNIA

Charles M. Campbell, Jr., M.D.
Consulting Pediatrician
Richard H. Lambert, M.D.
Consulting Psychiatrist
Ivan A. McGuire, M.D.
Consulting Psychiatrist
David L. Reeves, M.D.
Consulting Neurologist

Robert L. Brigden, Ph.D.
Director of the Ranch School
Thomas W. Jefferson, Ph.D.
Clinical Psychologist

Because

The Parents' Trust is in You

You should know the facts about the Devereux program of Education with Therapy

EXCELLENT professional facilities and long experience have equipped Devereux Schools to give effective education with therapy to the boy or girl whose emotional disturbances block his normal ability to learn.

Under the guidance of psychologists and psychiatrists, the staff develops a highly individualized program for each child. The group in which the child lives is chosen carefully to meet his social and educational needs, providing constant companionship and consistent stimulation towards natural family relationships.

You may have occasion to advise parents whose child may benefit from this program. The Devereux Schools staff will be pleased to study the case carefully and offer a detailed report on the possiblity of utilizing the Devereux program of education with therapy.

Please address your inquiries to
JOHN M. BARCLAY, *Registrar*, DEVON, PENN.

DEVEREUX FOUNDATION
CALIFORNIA PENNSYLVANIA

Devereux Schools

HELENA T. DEVEREUX, *Director*
J. CLIFFORD SCOTT, M.D., *Executive Director*
SANTA BARBARA, CALIFORNIA • DEVON, PENNSYLVANIA